

Half-way report on Chernobyl

The Soviet Union has published this week only the least interesting parts of its report on the Chernobyl disaster. More needs to be said.

HUMAN error is always a convenient explanation for catastrophe. It has the advantage, for the managers of great enterprises, that a few heads can be made to roll with some publicity without interfering too directly with the underlying programme. This is the spirit in which the surviving managers of the Soviet Union's still ambitious nuclear programme will no doubt be relieved by the publication earlier this week of an abridged version of the report on the Chernobyl disaster drawn up by the official investigating committee (see page 295). None of this implies that the explanation is incorrect. There is no reason to disbelieve the Soviet statement that the managers of the plant at Chernobyl behaved foolishly and "irresponsibly", or that they deserve to lose their jobs. Indeed, most Soviet statements about Chernobyl have been shown to have been accurate.

Nevertheless, the public interest in the Chernobyl disaster stems from the general concern that the world at large should have such a good understanding of how the accident happened that it is possible to make rational decisions about the future use of nuclear power as a source of electricity. Some, no doubt, seek further information so as to berate the hapless managers of nuclear industries elsewhere. Others seek the same information in the hope that it will demonstrate that nuclear power stations can be safely operated. The curiosity of both groups is legitimate, and must be satisfied. So must be that of the people most directly affected, the population of the region stretching for 400 square kilometres around the plant. The need now is for such a full disclosure of the circumstances leading to the accident that people will be able to judge for themselves whether the Chernobyl managers were as feckless as they are now said to be.

On present evidence, there is a long way to go before the circumstances are plain. It is said that those responsible at Chernobyl had been engaged in "experiments" not authorized by the licensing authorities when the accident happened, but there is no way of telling whether this was idle technological curiosity (unlikely on the eve of the May holiday) or, instead, a desperate attempt by unorthodox means to bring a wayward reactor under control. Similarly, there is not yet a sufficient explanation of the course of the accident itself. Monitoring stations elsewhere now clearly suggest that there was a second release of radioactive material from the reactor on the second day, presumably as a consequence of a further increase of the internal temperature, but the explanation remains for the time being hidden. Yet until this information is available, people outside the Soviet Union will remain on tenterhooks. There is again no reason to fear that the Soviet Union will not keep its promise to tell all. But this once secretive government now experimenting with openness clearly still has a lot to learn.

The Soviet managers of the nuclear industry need also to learn a little more caution about reactors of the type that went wrong at Chernobyl. One of the defects of this type of reactor is that the system becomes more reactive when water is lost from the cooling system, either because it leaks away or because it turns into steam. The consequence is an uncomfortable degree of instability which can also be a source of danger when something else goes wrong. But Soviet reactors of this type also lack the large pressure-tight containment vessels that would be required

of them in the United States and most other places. It must now be a matter of grave concern whether the other reactors of this type, of which are more than twenty still in operation in the Soviet Union, can be safely allowed to continue.

In one respect, the Soviet government has responded constructively to the delivery of the report by setting up a new ministry with responsibility for the design and construction of new reactors. Obviously the intention is that the Soviet nuclear industry should in future be run on lines more appropriate for a military organization (which, in the circumstances, may be the best solution, although one conflicting with Mr Gorbachev's belief that there are too many ministries as there are). But the Soviet authorities should also pay some attention to the possibility that there may be other, better, solutions. Notoriously, Soviet industry is perennially crippled by the bureaucratic environment in which it must function. Part (but only part) of the trouble is that managers are responsible both to the government agency or ministry which has built their plant and to the Communist Party, working through its local and regional committees. This is a recipe for the division and even dilution of responsibility. At Chernobyl, it seems to have been a recipe for disaster as well. Nobody would expect that a single nuclear accident would persuade the Soviet Union to forsake the path Lenin mapped out, but there is good reason why the Soviet Union should now be asking whether its present arrangements for the management of industry are the only means by which the dictatorship of the proletariat can be exercised.

The plight of those exposed to fallout near Chernobyl but not acutely injured remains to be determined. Strictly speaking, it is a matter for the Soviet government, in its dealings with its own people, to decide. This week's statement says that those affected by the enormous disruption caused by the accident will be compensated, though it is not clear what that means. But there is a great need that there should be close medical surveillance of those affected for several decades to come. It is not merely that there is a great deal to learn about the effects of low-level radiation exposure, or that there are ethical considerations which require that those now at greater risk from cancer than most others should be given as timely a warning as possible, but that this is a chance for the Soviet Union to make good its previously poor record for compassion. □

Packing for the summit?

There may be a Soviet-US summit this year if only a package of agreements can be found.

OPTIMISM earlier this year that 1986 would be the year in which the first arms control agreement this decade would be signed seems to be strengthening again, against the odds, and in spite of the succession of disagreements between the two major powers that have studded the past few months. And while the latest developments amount to nothing more substantial than straws in the wind, they do suggest that the two sides will eventually talk to each other, at a summit meeting later in the year. That is something for which many people will be grateful. But the likely



content of any agreement that may emerge is still a matter for conjecture. What is going on?

The straws in the wind are several. The Soviet Foreign Minister in London last week referred to the summit meeting as if it had already been arranged, apparently anticipating events. At the Stockholm meeting within the framework of the Helsinki agreements on European security, the Soviet Union has dropped its previous demands that air movements should be notified as part of the package of "confidence-building measures" on which the negotiators have been working since the beginning of 1985; now there is a chance of something to be signed at Vienna in November. Less formally, the Pugwash organization seems to have had a constructive meeting in Moscow earlier in the month, leaving many of the participants with the impression that the Soviet side is both eager for a test ban and willing to be flexible about the terms in which it might be drawn. Then President Reagan has made an encouraging speech, saying that there is much that might be done with the most recent set of proposals to the bilateral negotiations at Geneva, suggesting that an agreement may in due course emerge. Why all this sweetness and light?

The simple explanation is that both the Soviet and the US leaders wish there to be a summit, but that the Soviet side is determined that it will not allow itself this time to return empty-handed, as in many ways it did from last year's meeting at Geneva. And the practical question is whether there is a deal of some kind that would satisfy Mr Gorbachev without sticking in Mr Reagan's throat (or the throats of his colleagues in the US administration). In many ways, guessing what the package will contain is like a children's nursery game, except that it seems clear that there can be no substantial agreement on the main issues at Geneva — reductions of strategic weapons, regulation of intermediate-range missiles and the Strategic Defense Initiative (SDI). So what else? Agreement to ratify unratified treaties, such as the threshold test-ban treaty whose provisions require the exchange of seismic data relevant to the monitoring of tests, would be a sensible step forward — one that should have been taken several years ago. But the Soviet Union, with its unilateral moratorium on testing still in force after nine months, seems anxious to secure substantial improvement in this field. That is apparently why so much attention is being paid to the several variations on the test-ban theme suggested at an informal meeting in Washington earlier in the year — comprehensive test bans whose start is delayed, or which may be interrupted occasionally. It should not hurt either side to put together a package of proposals along these lines. The obvious snag is that such a package would, by definition, not touch the centre of the problem of arms control, the regulation of major strategic weapons already deployed.

That is why the two participants in the summit should be most of all concerned to reach a new understanding on the principles by which future arms negotiations will be conducted. Too much has changed in the past few years for most people's comfort. President Reagan, who came to office nearly six years ago with a long string of cogent complaints against the then existing agreements (SALT II in particular), but who found himself forced by events and pressure from his electorate to take arms control more seriously, has now further muddied the waters by suggesting that SDI seriously offers a way round the present strategic doctrines based on mutual deterrence by the threat of a catastrophic nuclear exchange. But the plain truth is that SDI will never be as effective as its enthusiasts claim, and will probably yield nothing more than a space-borne early warning system whose existence will not substantially change the present strategic balance (if that is what it is). So the summit will be a wasted opportunity if it does not yield an understanding of the ways in which SDI has changed (or, rather, left unchanged) the rules of arms control. For pride's sake, President Reagan would need such an understanding to be limited in time, but that should be no problem. □

Research directions

The British government plans to evaluate its research. But there are pitfalls.

NOT much has yet been heard from Mr Kenneth Baker, the successor in the British government to Sir Keith Joseph at the Department of Education and Science, about his plans for the administration of research. Perhaps inevitably, most of his time seems to have been spent on the urgent problems of how to restore contentment and achievement to the schools, whose bruising dispute with the government over pay, working conditions and career prospects has left deep scars. Yet Mr Baker cannot overlook the problems thrown up in the past few years by the government's insistence that the costs of reorganizing (or cutting) the research councils should be met from within a static budget. Already, the political calendar has reached the point at which final estimates for the next financial year are being drawn up. Moreover, unless Mr Baker is careful (and quick), he may find that this important part of his parish has been taken away from him.

That is one implication of the decision, announced at the end of last month, that there will be a unit within the Cabinet Office for the assessment of science, technology and the relationship between them. The plan is that there should be a small group of people, run by a senior (but not too senior) civil servant, with the task of throwing light on the relationship between research spending and such benefits as may ensue. The hope is that the government will then be better placed to decide which expenditures on research will yield the greater benefits. Although the unit will be chiefly concerned with what government departments at present spend on research on their own account, and may thus be a means by which some external appraisal of the civil value of defence research is at last attained, it is inevitable that the boundaries will be fuzzy, and that the new unit's opinions will overlap with those of other organizations, the Advisory Board for the Research Councils in particular. Given its place within the bureaucracy, the new unit could powerfully influence the pattern of all publicly sponsored research.

So will the unit come to sensible conclusions? Even the passage of time may not be enough to tell, for the new unit (which still lacks a head) will almost certainly operate behind closed doors. And because its existence springs from the British government's well-known impatience with the notion that there is no way of telling in advance what investments in research are likely to be profitable, there can be no assurance that the unit will recognize as impediments to its work the two serious objections to over-closely directed research: the circumstance that successful innovation requires that companies should be both technically well-equipped and well-placed in the market and the temptation (from which civil servants are not immune) to be swept up by fashion. Mr Baker's own first claim on public attention, his advocacy of information technology as a means not merely of making Britain prosperous but of curing unemployment at the same time, was plainly timely in the early years of this government, but has not produced the expected benefits for the simple reason that Britain lacks sufficient technical skill to make full use of the opportunities that abound in information technology. It is especially shocking that the University Grants Committee, in its recent set of circulars to British universities, should have had to explain that the annual intake of students to engineering schools cannot be increased more quickly for lack of suitable school-leavers.

That is but one illustration of the truth that there is more to the problem of winning benefit from research than the backing of good ideas with sufficient funds. The most obvious pitfall for the new Cabinet Office unit is that it will be blind to pitfalls such as these. The danger is especially great because the unit's work will not necessarily be public. Even at this late stage, there may be a case for changing that. □

Soviet inquiry

Chernobyl accident is blamed on human error

THE Chernobyl disaster was due to human factors or, in the words of the official dispatch, "a whole series of crude violations of the regulations for the operation of reactor installations". Such was the decision of a special session of the Politburo last weekend, after considering the report of the governmental commission that has been working at Chernobyl since the accident on 26 April. The full text of the report is expected to be published in mid-August.

According to the Politburo statement, during the night of 25/26 April, unauthorized investigations into the operating conditions of the turbogenerators were carried out on the Number 4 set, while it was out of service for a planned overhaul. Just what form these "experiments" took is unclear. The dispatch merely stated that the management and specialists of the power station had neither prepared themselves for the experiments nor coordinated them with the appropriate authorities, although they were legally bound to do so. Furthermore, there was no proper supervision of the experiments, nor were proper safety measures enforced while they were in progress. Several high officials have been dismissed for allowing this situation of "irresponsibility, negligence and lack of discipline" to develop.

By putting the blame squarely on the Chernobyl crew, the report would seem to exonerate the Soviet nuclear power programme and, in particular, the design of the RBMK reactor. But the unauthorized, unsupervised "experiments" could possibly have a more ominous explanation. In mid-May, some three weeks after the accident, there were some veiled references in the Soviet media to "experiments" at Chernobyl, followed by unattributed and unofficial reports that a malfunction on the Number 4 reactor had developed on the afternoon of 25 April, and that an emergency crew was still struggling to close it down when the explosion occurred. In this case, the "experiments" would refer to unauthorized or incorrect shutdown procedures, possibly undertaken as a last resort. In all events, the Politburo statement still fails to explain how an experiment with the turbogenerator could cause an explosion in the reactor.

Although the report was, presumably, completed only shortly before the Politburo meeting, the Soviet Procurator's Office has already instituted criminal proceedings against "the persons guilty of the accident at Chernobyl". In view of Mr Gorbachev's current drive for "open-

ness", the trial is likely to be accorded maximum publicity. The Politburo statement outlines the charges against the accused: 28 people dead, damage to the health of a large number of people, including 203 cases of radiation sickness so far, the cost of large-scale prophylactic work (including medical checks on several hundred thousand people), 1,000 km² of agricultural land taken out of production, the closure of various enterprises, the cost of evacuation and the rehousing of the evacuees, the loss of power production. The direct losses due to the accident are estimated at 2,000 thousand million roubles.

For the moment, the Politburo seems more concerned to ensure that such negligence will not occur again. Four senior officials have been dismissed: Evgenii Kulov, chairman of the state committee for safety in the nuclear power industry, G.A. Shasharin, a deputy minister of power and electrification, Aleksandr Meshkov, a first deputy minister of medium engineering, and Ivan Emelyanov, a deputy director of a research and development institute. In addition, all have been subjected to "rigorous Party penalties", as has also Bryukhanov, the former director of the Chernobyl power station who was dismissed in May.

The Minister for Power and Electrification, Anatolii Mayorets, was allowed to retain his post, but suffered a Party reprimand, and the public humiliation of being told that the Politburo had taken into account that he had only recently been appointed to the job, and that, "if he did not draw the appropriate conclusions", he would suffer a more severe penalty.

A new All-Union Ministry of Nuclear Energy has been established, in order to "raise the level of management and responsibility" in the nuclear power industry, and plans are under way for the retraining and requalification of nuclear power personnel. The use of simulators in training programmes is to be expanded, presumably to familiarize personnel with emergency procedures.

Furthermore, although the RBMK reactor design appears to have been officially exonerated, the Politburo has issued an urgent "demand" to all ministries and departments that they should draw up and implement additional measures to ensure the safety of existing reactors — and an appeal to scientists of the world to cooperate in developing a new generation of fission and, in the longer term, fusion reactors.

Vera Rich

US and Pakistan

Agreement on technology

Washington

THE United States and Pakistan have signed an agreement to establish a framework for trade in high technology goods. The agreement will make it easier for Pakistan to obtain mainframe computers and modern telecommunications equipment but will also provide specific assurance that Pakistan will neither divert the goods to other countries, nor use them in ways not approved by the United States.

Pakistan Foreign Minister Yaqub Khan and US Secretary of State George Shultz signed the agreement at the White House last week during the visit of Prime Minister Mohammad Khan Junejo. Last year the United States signed a similar memorandum of understanding with India during Mr Rajiv Gandhi's trip to Washington. The Indian agreement has far greater economic significance than that being worked out with Pakistan. Trade with India amounts to approximately \$1.600 million per year, according to Commerce Department figures, and something like \$75 million of that is in high technology equipment, including computers.

By contrast, trade with Pakistan amounts to only \$350–400 million per year, with most of that consisting of aircraft, soya bean oil and heavy industrial equipment. According to a Commerce Department official, Pakistan is just beginning to automate its government, and with no domestic computer industry it will have to import everything it needs.

But the real significance of the Pakistan agreement may be political. Ken Flamm, a research associate at the Brookings Institution in Washington, says that the trade agreement will smoothe the Pakistani feathers, ruffled by the lack of an arrangement such as that already offered to India. At the same time it will give the United States some control over the way in which the high technology equipment is used.

The United States is primarily worried that Pakistan might use computers obtained under the agreement to further a nuclear weapons programme. While the United States does not believe that Pakistan now possesses nuclear weapons, it takes the view that Pakistan could produce such weapons in a fairly short space of time. Prime Minister Junejo took some pains during his visit to emphasize the peaceful nature of Pakistan's nuclear programme. A joint statement issued by the two governments affirmed the Prime Minister's support of "US efforts to promote arms control and the non-proliferation of nuclear weapons". **Joseph Palca**

AIDS

Confused ethics of blood testing

WHILE the publication this month of the results of compulsory blood tests for the acquired immune deficiency syndrome (AIDS) virus in potential military recruits in the United States has provided valuable, and to some extent disturbing, information on the spread of the virus, monitoring attempts in Britain are foundering on ethical and bureaucratic rocks.

According to Sir Richard Doll, who has been asked by the British Medical Research Council to try to overcome the problems, all concerned are agreed that it would be unethical to take blood specifically for AIDS testing without informing those being tested of the purpose and result of the test. But the contentious question is whether the testing of blood samples taken for completely different purposes should be allowed.

Those who argue against such tests do so on the grounds that it would be unethical not to pass on the result of positive tests to the individuals concerned and yet unethical also to inform them because of the lack of prior consent. Those for whom the need to monitor the spread of the virus is paramount advocate that tests are carried out in an anonymous way so that it would be literally impossible to identify the individual that provided the blood sample. The justification for brushing the ethical problem under the carpet is that without information on the spread of the virus, it will be hard to persuade those at risk, including politicians, of the need for preventive measures.

As sexual transmission is the only way by which the virus can spread in the general population, it is necessary only to monitor blood samples from sexually active groups. One obvious source of such samples for anonymous testing would be antenatal clinics, but the Royal College of Obstetricians and Gynaecologists has objected to the idea. Another source being investigated is sexually-transmitted disease clinics, which have the advantage of including both sexes.

At present, the compulsory tests on potential blood donors provide the only source of information on the spread of the AIDS virus in the British population at large. Among 140,000 potential new donors tested between February and May this year, 7 individuals (or 0.05 per thousand) had a positive test. From October 1985, when the testing began, until May 1986, the rate was 0.02 per thousand but it is impossible to determine the rate per donor, according to Dr Harold Gunson, who collates the figures, because a substantial but unknown number of donors have given blood on more than one occasion during that period.

Prevalence among new donors is the

best figure to monitor for a trend but none is yet discernible, says Gunson, who is in the process of determining the sex and age distribution of the donor pool so that prevalences can also be worked out on that basis. The overall prevalence among UK blood donors is about 20-fold less than it was in August 1985 in the United States, the latest period for which figures are available.

The UK Ministry of Defence is opposed to the compulsory testing of potential military recruits. But in the United States that source has provided valuable figures on the prevalence of the virus in a sexually active heterosexual population.

Figures published in the 4 July *Morbidity and Mortality Weekly Report* show that the overall prevalence from October

1985 to March 1986 was 1.4 per thousand. Figures varied widely by region and age group, ranging from 0.2 to 10.1 per thousand. The latter figure was for those of 26 or more in the central states on the Atlantic seaboard. At last month's huge AIDS conference in Paris, a figure of 20 per thousand was given for 18–25 year old recruits from Manhattan. Just over 300,000 potential recruits have been tested overall, with 459 positive results. The prevalence rate in males was three times that of females.

Seropositive applicants are excluded from military service, illustrating the growing discrimination that discourages people from agreeing to be tested and physicians from the prospect of having to inform anyone who has not specifically agreed to be tested of a positive result. Exclusion from life insurance and the possibility of losing one's job are added discouragements.

Peter Newmark

Japan

Islanders put environment first

Tokyo

JAPAN'S Liberal Democratic Party (LDP) won a landslide victory throughout Japan in the double election on 6 July, except in Miyakejima, a tiny Pacific Island south of Tokyo, where the party suffered a crushing setback. Local and worldwide opposition from environmentalists to government plans to build a US military airfield on the island probably helped to sway the voters.

Miyakejima is an idyllic volcanic isle blessed with coral reefs, luxuriant tropical vegetation and two rare endemic species of bird, the Izu Island thrush (*Turdus celaenops*) and Ijima's willow warbler (*Phylloscopus ijimae*). But if military planners have their way, the island will become an "aircraft carrier" for US Navy warplanes to practice night landing.

The local islanders, however, have put up dogged resistance to the government's plan. All attempts by the Defense Agency to hold public hearings on the issue have failed and two island assemblymen have been threatened with recall for failing to voice their opposition to the airfield. In the general election the Japan Socialist Party won easily over LDP with a vote double that of the previous election. International support for the islanders' cause has grown rapidly and they can now count British royalty amongst their allies.

In late May, the Wild Bird Society of Japan reported that the Duke of Edinburgh, president of the World Wildlife Fund, had written to Prime Minister Yasuhiro Nakasone and the US Department of Defense appealing for the plan to be called off to protect wildlife in the island. Following this the International Council for Bird Preservation passed a

resolution in June opposing the airfield's construction.

The Environment Agency carried out a survey in the area of the planned construction after a volcanic eruption in 1983 and concluded that "artificial objects would destroy the area if put up in defiance with harmony of the scenery". But the agency has yet to clarify its official position.

Numerous LDP delegations have visited the island to try and woo the islanders with promises of investment. But to no avail. As one islander put it: "important people from the LDP say they will make the island a 'mountain of treasure' but the island is already a mountain of treasures".

David Swinbanks

More dead penguins

THE mysterious mortality of rockhopper and gentoo penguins observed in the Falkland Islands (see *Nature* 322, 4; 1986) has now spread to Argentina. But the Argentine National Atomic Energy Commission (CNEA) has ruled out the possibility, suggested at a press conference at the Soviet Embassy in Buenos Aires, that radioactivity is involved. According to the Soviets, the four British ships sunk during the 1982 Falklands conflict had been carrying nuclear weapons, the casings of which had leaked, contaminating the waters. The CNEA scientists, however, are preserving a proper academic detachment on the issue. Asked to comment on the deaths of some 300 penguins near Piero Deseado, they stated categorically that, even if some radioactive material had escaped into the sea, the concentration would have been far too low to be dangerous.

Vera Rich

Engineered organisms

Confusion over European rules

THE announcement this week of a new and supposedly safe vaccine against rabies (see pages 304 and 373) may bring to a head a simmering conflict in Europe over whether to permit the release of genetically engineered organisms.

The new vaccine is a version of the vaccinia virus, which is related to smallpox but much less dangerous, engineered to express a protein from the coat of the rabies virus, and thus create immunity in any creature inoculated with the vaccine. Target populations are foxes in Europe and raccoons on the east coast of the United States. Both, in principle, could be fed the vaccine in infected bait.

Developed by the French genetic engineering company Transgène in Strasbourg, the vaccine will be commercialized by Mérieux of Lyons, where chief scientist Philippe Desmettre said last week that there are no immediate plans for field trials. French regulations require approval of such trials by the ministries of health and agriculture, but Mérieux has not so far applied for permission. The company has supplied vaccine for tests by the Wistar Institute in Philadelphia, but use of the vaccine in the wild is likely to take place in Europe first, said Desmettre.

In Europe, there is no general agreement on how to regulate field trials. Denmark has jumped the gun and already imposed controls; in West Germany and

in the European Parliament, the Green Party is pressing for action to prevent the release of genetically engineered organisms, but in Britain, France, the Netherlands and elsewhere there seems little concern. In the midst of this confusion, the European Commission is attempting to create a little order rather than see yet another European market divided by multifarious regulations. But the Commission itself is a house divided, with those representing scientific, industrial and agricultural interests in conflict with the environment directorate.

Nevertheless, Commission officials said last week that a paper on the problem is now circulating and is likely to be presented to the European Council of Ministers towards the end of this month. This, however, would merely alert ministers that there is an issue to be discussed and it is not even clear at what council the issue would be thrashed out. The European Parliament, whose irritant value — if not real power — is steadily increasing in the European pantheon, will debate a hard-hitting report on the same issues by Dutch socialist Phil Viehoff, later this year. Not only will Viehoff present her report, but six other committees covering biotechnology and the environment will also table positions, leading to what could be an acrimonious debate.

The Commission itself is aiming to sub-

mit joint safety proposals to Council by April next year — Denmark's "banning" of the release of engineered organisms, hailed as a success by Europe's greens, is not a major obstacle, for trials are not truly "banned" although they must be referred to the government for approval.

In the next few weeks, a major report should appear by the Organisation for Economic Cooperation and Development (OECD) in Paris on "safety considerations in the use of recombinant DNA". It is said by those who have seen it to be a "very reassuring report", arguing that risks are small. Britain is expected to base its own guidelines on the OECD recommendations.

Robert Walgate

Chinese science

Back to your work units

CHINA is to introduce new rules on foreign study. Although the number of people sent abroad to study at government expense is to remain at its present level, the mix will be different. There will be more emphasis on those applied disciplines necessary for China's modernization programme, and less opportunity for students to go abroad immediately after completing their first degrees. On the other hand, the number of people with masters' degrees who go abroad to work for a doctorate will be increased. Students who go abroad at their own expense (mainly the children of top officials, plus the lucky few who have wealthy relatives abroad) will be given greater "guidance" to ensure that their chosen studies also conform with the state plan.

The changes are meant to ensure a sufficient supply of trained personnel to satisfy the demands of "work units" throughout the country. "Poaching" of graduates from their assigned jobs is now a problem in China. Understaffed enterprises frequently offer substantial and fringe benefits to well-trained recruits; they even advertise in newsletters which flourish in spite of official disapproval. The new rules should frustrate the "poachers". From this year onward, most of the quotas for government-sponsored students will be assigned to enterprises and work units in need of trained staff; the latter will then select who is to go abroad.

Foreign study, it is stressed, is not meant to be simply a means of self-betterment. Those who obtain a doctoral degree abroad will not now be permitted to stay on just to acquire a still higher qualification. Instead, they will have to return to an appropriate job in China for several years, before being sent abroad again. This, it is explained, will better help them coordinate their studies with China's needs.

Vera Rich

Space antiques hedge against inflation

Washington

"MUSEUM piece" is an apt term for a satellite the US Air Force plans to launch next October. Although it is now called the Polar Beacon Experiment and Auroral Research (Polar BEAR) satellite, for eight years it hung in the Smithsonian Institution's National Air and Space Museum in Washington as a Transit 5A, an example of an early navigational satellite.

The satellite's transformation came at the hands of the Johns Hopkins University Applied Physics Laboratory (APL) in Laurel, Maryland. APL, the contractor for Polar BEAR, realized that the satellite shared many components with the Transit 5A and Oscar satellites it had built in the 1960s and early 1970s for the Air Force. In 1975, the Smithsonian was looking for an example of an early navigational satellite to hang in its new museum. APL offered an Oscar 17, with flight-tested components, that was made to look like the earlier Transit 5A. To get its Oscar satellite back, APL offered the Air and Space Museum an untested back-up Transit 5A satellite that had been on display at APL. By using the older parts, a saving of about \$2 million was

achieved on the \$15 million bill for the satellite. Much of the saving came from the solar panels. All but one of the thousands of solar cells on the panels worked perfectly when APL tested them.

A Scout launch vehicle will carry the Polar BEAR satellite into a circular polar orbit from Vandenberg Air Force Base in California. The Air Force will use the satellite for experiments to improve communications over the poles. In addition, an Auroral Imaging Remote Sensor will study the aurora borealis.

This is not the first time that the Air and Space Museum has transferred an exhibition piece to active duty. The University of Iowa borrowed part of the attitude control system on the museum's Applications Technology Satellite for its Plasma Diagnostic Package that flew aboard Spacelab 2 on the shuttle last summer. A foot restraint system, once used on Skylab and donated by the National Aeronautics and Space Administration (NASA) to the museum was used by shuttle astronauts. NASA also borrowed Skylab's shower, but only for developmental studies for the manned Space Station.

Joseph Palca

OECD indicators

Science statistics prove slippery

RESULTS for 1983 and data for a few countries for 1984–85 suggest a “definite slowing down” of research and development spending in the industrialized countries, according to the latest “science and technology indicators” report* by the Organisation for Economic Cooperation and Development (OECD).

Some doubts remain about the figures. The rate at which data are accumulated, checked and reported by governments means that most data in the OECD tables end in 1983 and the latest trends are hard to spot. Between 1979 and 1983, research and development expenditure in the OECD states (most of the non-Communist industrial world from the United States to Japan) had been rising rapidly.

Other points made by the OECD study are that up to 1983:

- Research and development became

more concentrated in the largest countries.

- Japanese research and development grew faster than that in any other country.
- The United States remained the largest research performer, with the countries of the European Economic Community (EEC) second.
- Japan spent “substantially less” than the EEC countries in total.
- Research and development kept pace with or exceeded economic growth in most countries. Exceptions are Switzerland, Australia and the United Kingdom during 1981–83.
- Most OECD government funding went to defence and space programmes.
- Basic research received about 15 per cent of total OECD research and development funds in 1981.
- Industry increased its role in relation to government as a research and development spender, except in Switzerland.

The value of OECD’s figures on higher education research is questionable, however, according to a British research team which has just reported to the British Advisory Board to the Research Councils (ABRC). The report will not be published until the autumn, but its authors have indicated that OECD data are limited by their reliance on national government interpretations of the “Frascati agreement” of the 1960s, which was supposed to codify the reporting and comparison of national research statistics but has since lost its real effectiveness due to changes in institutional structures. Governments may also stretch the meanings of terms to suit their political convenience.

Some of the more glaring problems thrown up by the British work, which relates only to higher education, are in the varying definition of universities, and the attributions of government laboratory spending. Japan, apparently, attributes half of the total budgets (including salaries) of its 1,000 or so colleges, few of which are at full university level, to “research”; and the United States lumps the research of its big defence laboratories such as Lawrence Livermore and Los Alamos in with that of the universities. In the DM 20,000 million German university accounts DM 5,000 million is distributed in part as research, but turns out on analysis to be income from hospital beds, through certain medical universities’ private health schemes.

The most anomalous of the six countries analysed by the British group is Japan, but European countries as a whole also suffer in a major way by excluding all international research facilities (such as the Organisation for European Nuclear Research, CERN) from their higher educa-

tion statistics. (The United States, by contrast, includes its big science facilities such as Fermilab in its university figures.)

Altogether, the feeling in British science policy circles is that such anomalies allow the all-powerful Treasury to decry OECD figures which show declining research spending in Britain. That is why an independent, in-depth study was commissioned. It is believed to leave no escape from the conclusion that British research in the higher education sector has indeed been undergoing a relative decline — which is what, of course, everyone knew but could not prove. **Robert Walgate**

Academician’s Berlin arrest riddle

A PUZZLING diplomatic incident involving East and West Germany was resolved on Monday, when the West German authorities decided it would not be in the public interest to prefer charges against Professor Herbert Meissner, deputy general secretary of the East German Academy of Sciences.

Meissner had been on what the East Germans described as a routine business trip to West Berlin, travelling on a diplomatic passport, when he was arrested for shop-lifting. He was subsequently taken to Munich where he was questioned by BND, the West German intelligence service. He then, in some unexplained way, appeared in Bonn, at the East German permanent mission. The head of the mission, Herr Lothar Glienke, then lodged a strong protest at the West German Chancellor’s office.

According to Glienke, the charges against Meissner were totally fabricated, his journey to Munich was “abduction”, and his interview with BND had included the use of pressure and blackmail to try to make him betray his country. Glienke had therefore demanded an immediate guarantee from the chancellor’s office that Meissner would be able to return to East Germany immediately, and that his diplomatic passport and personal papers would be returned.

The West Germans, however, maintain that the shoplifting charge was genuine, and that instead of answering the questions of the West German police, Meissner had demanded to be taken to BND in Munich, saying that he wanted to defect. A spokesman for the West German government, Herr Friedhelm Ost, stated, moreover, that Meissner combined his duties for the academy with work for the East German Ministry of State Security.

Either explanation of Meissner’s movements last week leaves a number of questions unanswered. The West German decision to drop charges will, however, presumably bury these questions in diplomatic obscurity. **Vera Rich**

Getting tough on doctorates

How to make do with little might be thought of as the theme of the new corporate plan from Britain’s Economic and Social Research Council. Taking inflation into account, the £23.6 million budget is lower than it has been for a decade and ways have to be found to be more selective.

The major move is to take £2 million from the budget for postgraduate awards and shift it to research over the next five years. The shift is largely dictated by complaints over the appalling performance of doctoral students supported by the council. Only 20 per cent of students submit theses within four years and around half never write up their research at all.

The council is now to take a tough stand. Institutions where fewer than 10 per cent of students submit theses within four years will be banned from taking on new students for two years. By 1989, 60 per cent of students must submit theses before new funds will be awarded. No action is planned against individual students: supervisors are to be held responsible. One hundred studentships will go over the next two years.

More money is to go into research but the council is to tighten the reins. Expenditure is to be concentrated on “research initiatives” where the council determines what will be done rather than on research grants where researchers choose for themselves. While this gives some assurance of the quality of research, it does, as the plan points out, lessen the chances of throwing up “original ideas with high future value”. Economic and social change in contemporary Britain is the current key research area.

Alun Anderson

UK universities

Research gradings stir emotions

BRITISH university departments that received poor marks in the recent University Grants Committee (UGC)'s assessment of research may have a chance to be re-rated if they can do better. Sir Peter Swinnerton-Dyer, chairman of UGC, spoke of the need for machinery to pick up those that improve "dramatically" in giving evidence to the House of Lords Select Committee on Science and Technology last week. But no general reassessment is expected for 4–5 years and 15 per cent of the money available to the universities is now to be distributed according to the quality of research.

The assessment has provoked considerable outrage. Departmental groups have been rated as below average, average, above average or "starred" — of outstanding merit by international standards. But angry academics writing to *The Times* describe the assessment as "wholly lacking in intellectual credibility" and its methodology as "questionable", "extremely crude" and "seriously defective".

Many researchers even protest that they have been given no indication of how the ratings were arrived at. But the method was actually straightforward: "individual judgement helped by some factual records", as a member of UGC put it. Last year, a letter was sent to all universities requesting short (500–600 word) profiles of their departmental "cost centres" together with a list of five recent publications demonstrating the quality of research in each cost centre. Cost centres are larger units than departments: mathematics and chemistry have their own cost centres, but the biological sciences are divided into biochemistry and "other biological sciences", and the physical sciences into physics and "other physical sciences".

UGC's committees (for physics, biology, and so on), each with some 12–15 members then divided up into smaller subcommittees and parcelled out the replies. Further evidence was added: the number of grants and research studentships gained from the research councils over the past three years, the level of support from the large medical charities, the numbers of Royal Society fellowships held, "new blood" posts gained over the past three years and money gained from industry. Opinions were sought and meetings convened with the research councils — which gave their private rather than corporate views.

At the end of this first stage, cost centres were scored on a scale of 1 to 5; where there were many departments in one cost centre (zoology, genetics and biology are all "other biological sciences"), some subcommittees scored individual departments. Final adjustments were made at

meetings of whole committees where personal knowledge of outstanding researchers, or those that had slipped from their prime, counted for much. It was then decided not to mention specifically the below-average departments, nor to say how below average they were, but to issue just three ratings and the "stars".

Given the way the assessment was done, huge surprises cannot be expected; *en masse* the results confirm popular prejudice (see map). In the Celtic north and west and parts of the Midlands, below-average cost centres are in the majority, often by a very considerable margin. London (with a few exceptions), the south and north of England and Edinburgh and Glasgow are in the black. Towering above all are Oxford and Cambridge, the latter with a staggering 32 starred departments and none below average.

But it is not all so simple. Even universities with poor overall ratings may contain departments of excellence — Dundee with stars for biochemistry and applied mathematics for example. Do these indicate a few individuals' outstanding efforts? And there are whole universities which go against trends. Warwick had seven stars



In the hatched areas universities have less than one above average cost centre for each one below average; at the two black dots (Oxford and Cambridge) there are nineteen or more above average for each below average; in other areas the ratio varies from one to ten above average (University of Warwick) to below average.

and no below-average departments: does this suggest others might learn from its management?

Critics, on the other hand, would say that nothing useful can be gained from the assessment. The chief objection is that it was grossly unfair to look at cost centres rather than individual departments. Several different departments often had to share a single report and pool the five examples of their recent work, so that some researchers went virtually unrepresented. The problem is most severe for small arts departments where, as one historian put it, research on Ancient Greece, might be judged from a work on modern Europe.

Sir Peter Swinnerton-Dyer agrees that the assessment results have not been "perfect" here. Small departments may also have been penalized because there seems to have been little effort to look at productivity per staff member rather than a few examples of excellence. *Science Citation Index*, with weighting for quality of the journal cited, might have provided a database.

Another criticism is that the style of a university has not been taken into account. Salford, for example, scores rather poorly but has been held up by British Prime Minister Mrs Margaret Thatcher as a model university for its success in creating jobs. A cut in funding because it fails to reach certain research norms may be regarded as an implicit judgement of its aims. Other university departments simply feel they have been wholly misjudged — and can provide evidence that they should be rated differently. But no general appeal mechanism is planned.

Perhaps the most telling criticisms of the assessment are of the effect it may have. What made the assessment necessary is that funds need to be concentrated on research that will prove outstanding. A monolithic judgement may not do this but simply fulfil its own prophecies. Departments that have, in the past, been poorly funded by the research councils are likely to have gained poor ratings and will now find life even more difficult as they lose UGC funds. The judgement of UGC may make it harder for them to find funds from other sources too.

But, if it is original ideas that should be funded, it can be argued that the more independent sources of funds there are the better. How will poorly rated departments, among them those that contain a few excellent people, now get back on their feet? Sir Peter advised that universities should continue to fund the good and pick out for help those that can improve. He saw little to be gained by putting money into bad departments. Harsh advice perhaps, but the chairman of UGC knows, perhaps better than anyone, how little money there is to spare for university research.

Alun Anderson

SDI

UK scientists should take care

BRITISH scientists thinking of participating in Strategic Defense Initiative (SDI) research might do well to consider their legal position carefully. An incident in the United States in which basic research results were suddenly classified and what has recently become known about British Ministry of Defence (MoD) guidelines suggest that scientists may find it hard to protect their rights of free publication.

The terms of British involvement in SDI are enshrined in a memorandum of understanding between the British and United States governments. A research contract is placed by the SDI Organization (SDIO) in Washington, an office separate from the Pentagon, in one of five classified, military Project Areas, or in the \$100 million Innovative Science and Technology (IST) programme for basic science.

Exactly what regulations cover the latter programme are unclear. The SDI Participation Office in the MoD promises scientists the right to publish work that is also receiving SDI funding, with the provision that a copy of the paper of presentation is delivered to it and SDIO in Washington before publication.

Fundamental research will thus be free, as has often been claimed by the university liaison officer of the Ministry of Defence SDI office, George Gallagher Daggitt. But will this always be true? In response to a letter from Paul Labbett (8 April 1986), research contracts officer at Imperial College, London, Daggitt wrote: "It is however conceivable that exceptions to this rule could arise if there is the likelihood of disclosing operational capabilities and performance characteristics of developing military systems. In this case the contract for the work will clearly stipulate that responsibility for the release of information lies with the sponsoring office" (that is, SDIO in Washington).

The case of a US high-energy physicist, Dr Andrew Sessler, shows that successful SDI research can easily be classified. His team's work on the free electron laser was guaranteed open for publication until he showed how gigawatt power outputs were possible in March 1985. SDIO immediately classified the entire experiment and the results. The work, largely paid for by the Department of Energy, was part of the civil fusion reactor programme. But as Sessler took a small grant from SDIO it became the sponsor. The work was classified top secret (Sessler and his staff were threatened with expulsion and jail if they released details) for 13 months until April this year, when the results alone were de-classified. The experimental details are still classified.

SDIO is free to classify what it sees fit. The memorandum can offer no guaran-

tees that fundamental science will be open for publication. MoD Guidelines say a review procedure is available "so that appropriate representations can be made". But SDIO in Washington has the final say. A significant number of British scientists have applied for funding under IST: there are 36 individual proposals, and 2 consortia.

The memorandum embodies the principle that there will be no guaranteed work although the figure of \$1,500 million was floated early on. Days before the memorandum of understanding was signed, the British draft contained a schedule of work in three categories: commercial work from research into manufacture worth \$1,000 million, government contracts for research worth \$250 million and university research up to \$250 million. How much will British companies and individual researchers actually gain in SDI? From \$50 to \$100 million in total, estimates MoD.

MoD can identify five commercial contracts worth a total of \$1 million. Another 12 contracts are in negotiation, each worth around \$2 million. The first round of IST proposals might bring another \$10 million. Government-to-government business could rise from the present \$15 million to \$50 million.

The government agreed a special research contract for the work in which it is the principal or subcontractor. The Letter of Offer and Acceptance negoti-

ated to cover the first two contracts, "written according to patent law" according to the SDI Office in the MoD, acts as a standard contract for SDI research and is said to safeguard the rights of ownership of technology. The letter is secret and applies only to government work. Yet it is said to embody the best definition of intellectual property right, and to safeguard technology produced in SDI for the researcher concerned.

Why should it be secret? And why cannot other researchers use the definition of intellectual property right? MoD says it is for private individuals and companies to get the best deal they can for themselves even though the memorandum of understanding was intended to act as an umbrella agreement for all participants. The Americans insisted on this arrangement.

Government work can only be research, not manufacture, and must stay within the terms of the 1972 Anti-Ballistic Missile (ABM) Treaty according to the memorandum of understanding.

A joint MoD/Foreign Office Arms Control Unit has a veto on government research proposals that threaten to stray beyond other treaty obligations. If a British company were to design, prototype, test or build a component of SDI, however, it is unclear in the memorandum whether or not such proposals could be vetoed.

The right to own technology originating in a research contract funded either by SDIO (as basic science), or in a subcontract with a US company (on project work), is not defined in the memorandum. This is thought to have contributed to the

UK star wars consortium launched

A CONSORTIUM of eleven UK defence contractors has bid jointly for the first Strategic Defense Initiative (SDI) contract worth \$9.9 million, the European Architecture Study, let to the Ministry of Defence by the Strategic Defense Initiative Organization (SDIO) in Washington, DC.

Ian Sutherland, a director of Marconi, said that this "team" will share out research work if it secures the entire study and would work together as "UK Ltd" on the SDI contracts that follow on. These could be worth tens of millions of dollars.

The group includes defence contractors GEC-Marconi, British Aerospace, Barr & Stroud and Shorts, electronics companies Thorn EMI, Racal and Ferranti, the Royal Ordnance Factory, and computer software houses CAP Scientific, Scicon and Software Sciences.

The architecture study will consider how an SDI system could operate in Europe. Four subcontracts will be awarded, probably in September, in computer systems, weapons, battle management and command, control and communication.

Few contracts have yet been awarded to British companies. Sutherland said Marconi has sold "£4 million worth of hardware", including exotic electronic components such as thyristors (which switch very high currents quickly) and silicon-on-sapphire semiconductors (which are relatively unaffected by radiation). In addition, Marconi Projects has two research subcontracts worth under \$250,000.

Up to 12 other proposals are in the pipeline, including those for electromagnetic guns, battle management, lethality and target hardening, very high speed integrated circuits, high-energy lasers, advanced sensors and particle beams.

SDIO in Washington has also requested access to the Ministry of Defence's results from the Teal Ruby experiment which considered strategic defence from short-range and mid-range missile attack. But the Australian government, which collaborated in the experiment, is set against SDI participation and Australian permission is necessary before results can be released.

Paul Walton

low take-up of work by companies. Background research, that is existing company data on intellectual property rights defined by patent or copyright, can be protected. But foreground research, that produced during a funded contract, is not protected by the memorandum.

As things stand, a US prime contractor would be legally entitled (under US law) to patent foreground research, denying use to the originator for 17 years. This patent defines the owner as the body which pays, not the individual or organization which completes the research.

A further problem for British researchers is that the memorandum does not waive those items of US law, in particular the Export Administration Act, that make



At least we've proved the existence of Black Holes...

it difficult for items of technology or technical data in the form of proposals or descriptions to be passed freely between Britain and the United States.

Technology transfer law still applies to SDI work. "Both Governments will make every effort to process within 30 days requests for export licences for technical data packages or other controlled information for direct bidding", according to the memorandum. In practice, clearance can take several months, by which time an SDI contract will have been let. Visits to secure sites, to closed conferences, or to discuss classified information with a partner in the United States requires prior clearance, and full vetting, of any person by the US authorities. It is likely that positive vetting will be carried out by US embassy staff in the United Kingdom. Most British citizens will be required to sign the Official Secrets Act.

The memorandum is "secret in perpetuity", or top secret. A full debate on British participation has been requested by Labour MP Tam Dalyell but denied on the basis that the United Kingdom has negotiated a good deal which the United States does not wish other allies to know of in detail. All the UK opposition parties, Labour, Liberal and Social Democrats, have pledged that they will do away with the memorandum and pull out of SDI.

Paul Walton

US space

Shuttle faces more delay

Washington

THE US National Aeronautics and Space Administration (NASA) announced last week that shuttle flights will not now resume before 1988, despite earlier estimates that the shuttle might fly again as soon as next summer. Although disappointing, the delay, say NASA officials, is necessary to complete a review of the design and testing of the solid rocket motors (SRMs) responsible for last January's shuttle accident.

In a report requested by President Reagan, NASA spelled out its plans for implementing the recommendation of the Rogers Commission on the accident (see *Nature* 321, 637; 1986). NASA hopes to redesign the SRMs so that existing hardware can be used, but there are alternatives using completely new hardware in case that plan is frustrated. At NASA's request, the National Research Council (NRC) has established an independent oversight group, chaired by H. Guyford Stever, to superintend the redesign.

Other hardware modifications recommended by the Rogers Commission included improvements in the tyre, brake and nose-wheel steering systems. NASA says that some of those improvements were under way at the time of the accident, and that the other modifications are in hand. Until the improvements are judged a success, the shuttle will continue to land at Edwards Air Force Base in California, where there is a longer landing strip. Ultimately NASA hopes to land shuttles at the Kennedy Space Center in Florida where they are launched. NASA also plans a thorough review of all critical safety items on the shuttle, and a second NRC oversight panel is being formed to watch over that process as well. Not sur-

prisingly, safety issues are uppermost in the minds of most NASA officials. Waivers allowing marginal parts to be used in launches will be few and far between, says one NASA official.

NASA has also begun two separate reviews of its much criticized management practices. One will look at the shuttle programme, the other at NASA as a whole. Many at NASA involved in the shuttle have either left the agency or been transferred since the accident. The latest casualty is Lawrence Mulloy, director of the SRM programme at the beginning of the year, who resigned last week after having been transferred from that programme.

NASA has yet to make a final decision about crew escape systems, but preliminary conclusions suggest that no approach will provide a safe escape route under all conditions. But further attention is to be paid to procedures for aborting launches, including one scheme for transatlantic escape.

The Rogers Commission pointed out that total reliance on the shuttle for launches had put heavy pressure on NASA to increase its launch rate. NASA hopes that the pressure will be partly reduced by the new policies it has adopted on the choice of cargo for the shuttle bay, but it is also now backing the claims of other agencies that there should be a mixed fleet of launch vehicles, including expendable rockets. Responding to a request from Congress, NASA has asked NRC to form a third panel that will evaluate launch rates and the balance between manned and unmanned launch systems. Edward David, president of Exxon Research and Engineering Company, will chair this committee.

Joseph Palca

Now it's save *French* science!

FRENCH scientists appear to be getting more and more like their British colleagues — absolutely desperate. But in one way the French are going further. Discovering that appeals to their own government over recent cuts have failed (no doubt in part because, as it is said, the new science minister is uninterested in politics), they have gone to the lengths of producing an international petition: "Save CNRS" (the principal French research council).

"We the undersigned members of the international scientific community", the petition reads, "wish to express our grave concern regarding the policy followed by the French government. The policy has already resulted in: budget cuts (more than FF 4,000 million — £400 million); 25 per cent reductions in new posts for 1986; a

decision to cut the number of scientists employed by government in 1987; the smallest ratio of research spending to gross national product in the West; and the suppression of the CNRS Comité National." The Comité is a decision-making body whose loss brings to a standstill the whole administrative machinery of the CNRS — including the setting up of new posts and grants.

The petition calls for the French government to reinstate the Comité National, and reverse the new downward trend in jobs and cash for French science. What it fails to do, however, is to give an address to which to send the completed petition. But interested readers could always try the French Prime Minister, Jacques Chirac, at the Matignon... Robert Walgate

European research priorities

SIR—The Advisory Subgroup on Human Reproduction of the European Medical Research Council (EMRC)* has published recommendations for research related to *in vitro* fertilization and embryo transfer¹ and has sought to identify priority areas in human reproduction research deemed particularly suitable for investigation by European scientists².

The subgroup has noted the decrease of financial support for human reproduction research, particularly in the development of new and safe methods for fertility regulation, from both the public and private sectors³. This contrasts strikingly with the recommendations for increased support for reproduction research by all sectors called for by 146 governments at the International Population conference in Mexico City in 1984, where particular emphasis was placed on the special needs of developing countries⁴.

The subgroup believes that the scientific community has a particular contribution to make to research on new methods for fertility regulation for three reasons: its well-established record in this field, its long-standing tradition of providing medical and scientific assistance to developing countries and the close collaboration between its medical research councils and the European pharmaceutical industry.

In order to facilitate this collaboration, the subgroup held a consultation with representatives of interested European pharmaceutical companies in October 1985. At a meeting of the subgroup alone in April 1986, the following recommendations, which represent the views of the subgroup and do not necessarily reflect the views of the industry participants in the earlier consultation, were drawn up.

- (1) Steps should be taken, separately and jointly, by both industry and the medical research councils to increase the priority given to research on new methods of fertility regulation and on related fundamental research.
- (2) The requirements for clinical trials of new contraceptive drugs and for the approval of such drugs for public use vary from country to country and are deemed unnecessarily complex in certain instances. The subgroup recommends that these requirements be modernized and standardized.
- (3) A marked increase in the number and magnitude of liability lawsuits has caused some companies to withdraw from the fertility regulation field, and is one of the primary reasons for decreased work in this field. The subgroup recommends that European MRCs and industry representatives should work together to forestall this problem which threatens progress in this field.
- (4) In view of the challenges confronting

science to develop new methods of fertility regulation, the subgroup recommends consideration of the establishment of a European funding mechanism for reproduction research, sponsored and funded by industry and the MRCs and dedicated to the support of European research and research training adapted to the jointly derived requirements of both the private and public sectors.

E. NIESCHLAG
(Chairman of the EMRC
Advisory Subgroup)

MPG Clinical Research Unit
for Reproductive Medicine,
Steinfurter Str. 107,
4400 Münster, FRG

*The advisory subgroup includes: E. Nieschlag (chairman), Max Planck Clinical Research Unit for Reproductive Medicine, University of Münster; B.M. Baccetti, Institute of Zoology, University of Siena; L. Cedard, Maternité Cochin, Paris; P. Corfman, Special Programme of Research, Development and Research Training in Human Reproduction, World Health Organization, Geneva (observer); F. Haseltine, National Institute of Child Health and Human Development, Bethesda (observer); E.D.B. Johansson, Pharmacia AB, Uppsala; D.W. Lincoln, MRC Reproductive Biology Unit, Edinburgh; C. Robyn, Hôpital Universitaire Saint Pierre, Brussels; R. Vihko, Department of Clinical Chemistry, University of Oulu; G.H. Zeilmaker, Department of Endocrinology, Growth and Reproduction, Erasmus University, Rotterdam.

1. *Lancet*, ii, 1187 (1983).
2. *Lancet*, i, 1228 (1984).
3. Atkinson, L.E., Lincoln R., Forrest J.D. *Family Planning perspectives*, 17, 196 (1985).
4. *Report of the International Conference on Population*, United Nations Publication E/CONF. 76/19, New York (1984).

Henderson Island

SIR—In 1983 there was international concern when a strip-miner from West Virginia proposed to establish a settlement on Henderson Island in the Pitcairn group of the central South Pacific, as it was thought to be one of the few raised atolls still comparatively unaffected by human activity¹. As much of the case for preserving the island depends upon its undisturbed state, we are now concerned to observe a statement in an influential journal that: "The recent discovery of several Polynesian occupational sites on Henderson² shows that the island had, in fact, been inhabited in prehistoric times. On the basis of vertebrate remains from one of these sites we can now show that this period of human occupancy was accompanied by extinction and a consequent decrease in the species diversity of the island... The biota of Henderson Island can thus no longer be regarded as being in an unaltered state."³

As this statement is liable to be quoted in support of further proposals for development of Henderson Island, it deserves some scrutiny.

The past Polynesian presence, which was already noted in our recent contribution on the birds of the island⁴, had so little impact that it is difficult to detect. Its only observed effects appear to have been the

introduction of the Polynesian rat *Rattus exulans*, and perhaps some plants, and the possible extermination of some widespread but unobtrusive seabirds, whose survival may have been overlooked, and pigeons of the genus *Ducula*, attributed to two species because the wings were comparatively small whereas the jaw and legs were rather large.

The alternative possibility that instead of two additional species of pigeon occurring on one very small island alongside the small one which still survives, the pigeon bones may have belonged to a single large species showing the reduction in the wings common in insular birds, and also found in the flightless Henderson rail *Porzana atra*, which will be the next species at risk if the island is developed, is not discussed.

The list of petrels of the genus *Pterodroma* reported to survive in the group⁵ does not agree with our experience⁶, and it is rather surprising that all the bones should be identified as *P. alba* when it is doubtful whether this is distinguishable osteologically from one of the other species which is now commoner, *P. arminjoniana*.

It is to be hoped that this contribution will not be allowed to count against the case for preserving the island, which is apparently still at risk.

W.R.P. BOURNE

Department of Zoology,
The University,
Aberdeen, UK

A.C.F. DAVID

Oak End, West Monkton,
Taunton, Somerset, UK

1. Fosberg, F.R., Sachet, M.-H. & Stoddart, D.R. *Atoll Res. Bull.* 272, 1–47. (1983).
2. Sinoto, Y. S. J. *Soc. Ocean.* 39, 57–67 (1983).
3. Steadman, D.W. & Olson, S.L. *Proc. natn. Acad. Sci. U.S.A.* 82, 6191–6195 (1985).
4. Bourne, W.R.P. & David, A.C.F. *Notornis* 30, 233–252 (1983).

More bright ideas

SIR—We read with interest of Greg Ojakangas's award-winning proposal to use a ROOPH (Readily Operative Overhead Protection by Hippos) in the event of nuclear attack (*Nature* 321, 373; 1986).

We wish to suggest the use of BOOBS (Blockade of Overhead Objects by Baby Seals) based on the same principle.

Canada could supply the animals (having an abundance of baby seals, especially in recent years) in return for protection. Use of an indigenous species has strategic and financial advantages, and, unlike hippos, baby seals are already cold-adapted for deployment at high altitudes or in winter (nuclear or otherwise).

MARK RAGAN

KATHY SCHWARTZENTRUBER

1141 Oxford Street,
Halifax,
Nova Scotia,
Canada B3H 3Z1

Gentle warning on fractal fashions

The novel idea that fractal structures can support high-energy lattice vibrations may not be the philosophers' stone that some have suggested.

FRactal structures are understandably one of the captivating fashions of our times. Even without the proselytizing of B.b. Mandelbrot (as in *The Fractal Geometry of Nature* (Freeman, New York; 1983)), the subject would by now have been dragged into prominence by the host of practical applications to which it lends itself. The simplest problems are those of the aggregation of particles onto a growing aggregate, where there is now a host of demonstrations that the density of such a structure is a decreasing function of its size, most graphically described by means of a "dimension" smaller than that of the space in which it grows. But it is plain that biological applications must also abound. For are not large polymer molecules put together by the aggregation of monomers onto a growing backbone candidate fractal structures?

This line of enquiry has already been productive. Some polymer structures do indeed turn out to be fractal structures, while the suspicion has grown up that the properties of some important biological molecules are in many ways a consequence of their supposed inherent fractal character. But fashions that are too pervasive can also be misleading, and require occasional correction. That is the chief interest in an elegant paper by J.A. Krumhansl (*Phys. Rev. Lett* **56**, 2696; 1986) of Cornell University which, in the gentlest manner, warns people against the tendency to look for fractal explanations everywhere.

Krumhansl's topic is also interesting and potentially important. His starting-point is a series of speculations about the interaction in molecules such as those of haemoglobin of the central iron atoms with the surrounding ring of porphyrin rings and the protein structures to which they are in turn attached. Changing the spin states of the central iron atoms requires a substantial amount of energy, greater than that corresponding to the mostly low-energy vibrations of the usual protein chain. So how can a larger amount of energy be exchanged with a protein environment? Fashionably, there seems recently to have been a tendency to invoke fractal phenomena of a new kind, those pertaining to the vibration of fractal structures of various kinds.

The mere idea that it might be possible to calculate the vibration frequencies of a fractal structure is naturally surprising. To be sure, it should in principle be possible

to deal with any specific fractal structure provided there is enough computer power to hand. One common model is that in which the particles in an aggregate are replaced by masses (all of the same size) which are then supposedly connected to their neighbours with springs with identical elastic properties. Since most simulations of fractal structures begin life on an underlying lattice with well-defined properties, solving the vibrational problem is essentially a calculation of the vibrations of a lattice in which there are patches missing. And even though each different arrangement of patches should in principle give rise to a distinctive set of vibration frequencies, the fact that the geometrical properties of aggregates can be described in gross terms (by their fractal geometry) means that it is reasonable to look for general properties of the vibrational spectrum of a fractal lattice.

It is nevertheless remarkable that so much has been accomplished in just over a decade. The frequency spectrum of a reasonably well-behaved solid (with three space dimension) is a function proportional to the square of the frequency. So much has been known since Debye and others took up the question of the specific heat of solids early in this century. At low frequencies, fractal structures have the same kind of vibrational spectrum; longer waves do not recognize that there are patches of the lattice missing.

The surprise (but a little thought will show that it is not really that surprising) is that the vibrational spectrum of a fractal structure includes a large proportion of high-frequency vibrations. Crudely, there must be within a fractal structure patches filled with particles which are nearly isolated from the rest of the structure, and which then perforce vibrate internally at lower frequencies because they are denied the opportunity for taking part in the large-scale long-wavelength vibrations involving the operation of the lattice as a whole.

This conclusion obviously neatly fills the bill required by the haemoglobin problems. If fractal structures not merely admit but require an unexpectedly large proportion of vibrations at high frequency (or, more strictly, short wavelength), there will be opportunities for the exchange of relatively large amounts of energy between, say, haem iron atoms and the surrounding haemoglobin molecule. But there is one snag: the peculiarly

fractal vibrations are also localized (to the patches of the lattice which are themselves almost geometrically isolated), so that the rate of energy exchange will not be great. Is it possible to circumvent that difficulty?

Krumhansl's paper is not so much a protest as a demurring. He does not question the possibility that fractal structures may sustain short wavelength vibrations (which, being cousins of more familiar photons, have inevitably been called fractons), but he does insist that the statement cannot be turned around in such a way that every solid with an anomalous vibration spectrum can be inferred to be a fractal structure of some kind.

The issue is of some importance because of the way in which estimates of the vibrational spectra of real solids (most simply made by measuring the specific heat as a function of temperature) have been used to make structural inferences. This has become especially fashionable in the study of amorphous solids, often said to be fractal over short distances. For starters, Krumhansl offers the example of the monatomic indubitably three-dimensional solid selenium, whose vibrational spectrum is as anomalous as it could be.

Krumhansl's explanation is both simpler and more persuasive. The anomalous vibrational frequency-spectra are often merely consequences of anisotropy of the forces between the atoms in a vibrating structure. Where proteins are concerned, for example, the forces are not central but are directed along the chemical bonds by which atoms are joined. And, in other circumstances, bond-bending forces become dominant.

The effect of Krumhansl's gentle broadside will not be to discourage interest in fractal structures. Quite possibly, the effect may be the reverse of that (no bad thing in itself). But this is an interesting case, which seems to happen more and more often, of how an explanation in search of a phenomenon may lay claim to far more territory than it can handle.

For the time being, and perhaps for a long time to come, progress in the field of fractals hangs crucially on numerical simulation. The results are convincing but not suggestive. They tend to confirm conjectures but not to suggest them. One day, it may all be different. Let us hope so.

John Maddox

Rabies control

Vaccination of wildlife reservoirs

from Roy M. Anderson

CERTAIN major infectious diseases of man and his livestock persist endemically in so-called reservoir populations of wild mammals. In Europe, two important examples are those of rabies and bovine tuberculosis. Such infections are usually difficult to control and past efforts have often centred on the widescale slaughter or culling of the reservoir host species in the absence of more effective measures. But in this issue (Blancou, J. *et al.* *Nature* 322, 373; 1986) the production of a recombinant vaccinia virus is described that may be an important step towards the control of rabies by vaccination.

Rabies, an acute viral infection of the central nervous system, has throughout human history been one of the most feared of all diseases because of the distressing clinical symptoms it induces and their invariably fatal outcome. The current epidemic in Europe (see figure) is characterized by a high incidence in the red fox (*Vulpes vulpes*), which accounts for more than 70 per cent of all reported cases, and populations of this species are believed to constitute the principal reservoir of infection. Tuberculosis in cattle (*Mycobacterium bovis*) is of less cause for concern to man but the infection can result in serious economic losses to the beef and dairy industries and it can present a threat to public health in the absence of milk pasteurization. In many regions of Western Europe populations of the badger (*Meles meles*) form an important wildlife reservoir of tuberculosis.

In both examples difficulties have been encountered in the design and implementation of effective control measures. A significant factor is the widespread public concern at the slaughter of wildlife for the control of diseases that are of limited significance to human health and welfare. Human deaths resulting from rabies in Europe, for example, have been of the order of 1–4 per annum over the past two decades. In contrast, crude estimates suggest that a minimum of 1¼ million foxes are killed annually by rabies control programmes in Europe (Zimen, E. *The Red Fox*, Junk, The Hague, 1980). Economic as well as conservation issues are also of relevance. A recent report commissioned by the British government suggests that the cost of badger culling to reduce the incidence of bovine tuberculosis in cattle herds far outweighs the economic benefit of control (*Badgers and bovine tuberculosis*, HMSO, 1985).

The major problem, however, concerns the limited success achieved by wildlife culling. The extensive slaughter of foxes in countries such as France and West Germany has slowed the rate but not stopped the spread of rabies across mainland Europe. Similarly, efforts to remove social groups of badgers infected with bovine tuberculosis in areas of England and Wales have reduced the incidence of cattle-herd infection but they have not eliminated the problem. These experiences have stimulated a search for alternative or supplementary methods of

control and, in the cases of rabies, much attention has been focused on the development of safe vaccines for use in wildlife populations.

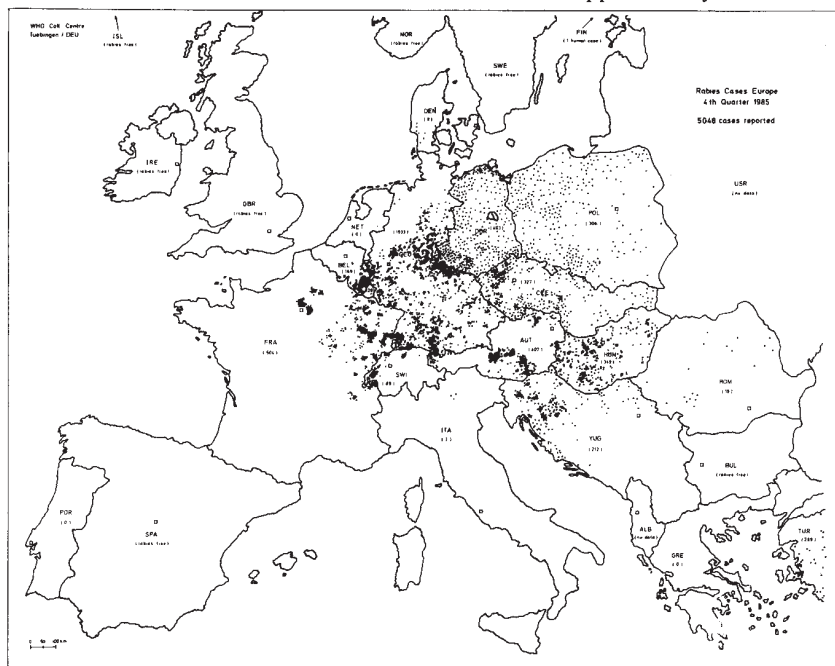
It is in this context that the work of Blancou *et al.* is important. Vaccines based on inactivated or attenuated rabies virus have been available for some time but their use to immunize wildlife has been restricted by poor efficacy and fears over their safety. Wildlife is vaccinated by consumption of inoculated baits (chicken heads or sausages) but this oral route of immunization for inactivated vaccines fails to induce strong protection. Oral administration of attenuated vaccines is better and appears to create effective immunity but the instability of the attenuated strains and the possibility of reversion to full virulence have prevented their use in natural habitats. Blancou and co-workers report the successful development and testing in foxes of a recombinant vaccinia virus harbouring the gene encoding the rabies surface antigen. Oral and intradermal inoculation give good protection against challenge with the wild rabies virus. The safety, efficacy and ease of administration of this new vaccine should soon lead to large-scale field trials in wild populations of foxes.

The development of a safe and effective vaccine, although essential, is only a first step in the control of infection within a population. The creation of sufficient herd immunity to block transmission is dependent on various economic and other factors, the biology and demography of the target host species and the epidemiology of the virus. A central question in the implementation of mass vaccination is what proportion of the target population must be immunized to halt disease spread. A related issue is the relationship of this critical proportion with factors such as fox density and spatial distribution.

A body of epidemiological theory is available today to help answer these questions. Simple mathematical models of the transmission of rabies in fox populations suggest that the critical proportion of animals, p , that must be effectively immunized to eradicate the infection is approximately given by

$$p > 1 - K_r/K$$

Here, K_r is the critical density of foxes necessary for the endemic persistence of rabies and K is the carrying capacity of the habitat or the magnitude of fox abundance in the absence of rabies (Anderson, R.M. *et al.* *Nature* 289, 765; 1981). Epidemiological observations on rabies persistence in different habitats that support varying densities of foxes suggest that the value of K_r is approximately 0.3–0.5 foxes km⁻². The value of K is determined by the mixture of habitat types in an area and within Europe it typically lies in the range of 0.1 to 4 foxes km⁻² (Harris, S. & Rayner, J.M.V. *J. anim. Ecol.* 55, 575; 1986). In



some urban and suburban areas fox density may be much higher (7–10 foxes km^{-2}). The equation shows clearly that the required degree of vaccine-induced herd immunity is greater in dense fox populations than in sparse ones; for $K \approx 2$ foxes km^{-2} the required degree of herd immunity is predicted to be approximately 80 per cent ($K_r = 0.4 \text{ km}^{-2}$). In urban areas with dense populations for example, ($K = 7 \text{ km}^{-2}$) the required level of protection is predicted to be roughly 95 per cent. These figures are depressingly high and question the practical feasibility of mass vaccination as a means of control.

A further problem is created by the demography of the reservoir host species. The fox has a high fecundity (an average birth rate of 4–5 cubs per litter each year) and a short life expectancy (between 1.5–2.5 years from birth). These observations imply that a high proportion of fox populations are young animals in their first or second year of life. As such the maintenance of high levels of herd immunity would require repeated (preferably annual) cohort immunization with the aim of exposing young foxes to inoculated baits at as young an age as is practically possible.

The development of a safe wildlife vaccine as described by Blancou *et al.* is

clearly an important step towards the control of rabies. In practice it will undoubtedly be of value to block transmission in areas of low fox density (given that the vaccine can be produced and delivered cheaply). But in areas of high fox density, such as many urban areas in Europe (where the risk of transmission to man and his domestic animals is greatest), mass vaccination will probably have to be supplemented by culling to reduce the target level of immunization (the degree of herd immunity) required to prevent rabies persistence.

The new vaccine's greatest value may well be in the immunization of domestic animals and pets as opposed to wildlife reservoir species. In Latin America, Asia and Africa, for example, rabies in dogs presents the major threat to man (most of the estimated 15,000 annual human deaths caused by rabies occur in these regions). In densely populated urban areas, with large free-roaming dog and cat populations, baits inoculated with a safe recombinant vaccine is potentially a very attractive method of reducing the risk of transmission to man. □

Roy M. Anderson is Professor of Parasite Ecology in the Department of Pure and Applied Biology at Imperial College, University of London, London SW7 2BB, UK.



gases, only one of which (occupying about a quarter of the volume) attracts phlogiston. This latter gas he termed *Feuerluft* (fire air) and eventually isolated it by heating concentrated nitric acid with charcoal, absorbing in milk of lime the red fumes of nitrogen dioxide generated. The oxygen remaining was formed by the thermal decomposition of the acid.

Scheele came close to anticipating Lavoisier by considering the possibility that during combustion fire air combines with the burning body, but he eventually concluded that the gas unites with phlogiston forming heat — which, like most of his contemporaries, he regarded as a material. This hypothesis enabled Scheele to explain why, in a few cases, a calx could be restored to a metal by heat alone, whereas normally the presence of a substance rich in phlogiston (for example, charcoal) was necessary.

Although misled by his belief in phlogiston, Scheele's record of achievement during his short life is remarkable. He discovered radiant heat, noting the similarity of its properties to those of light. His investigation of fluorite, pyrolusite and molybdenite led, respectively, to the discoveries of hydrofluoric acid, chlorine and molybdenum. He discovered glycerol and isolated a dozen previously unknown organic acids. A minor mystery surrounds his discovery of the toxic hydrocyanic acid, resulting from his investigation of the pigment Prussian blue, for he described its taste and smell.

From the age of fifteen Scheele worked in various pharmacies in Sweden until, in the summer of 1775, he acquired his own pharmacy from the widow of Herman Pohl in the small town of Köping. Though subsequently offered academic posts he refused to leave his practice. A few days before his death in 1786 he married Pohl's widow and made her his heiress. □

E.L. Scott is a retired schoolmaster and historian at Spring Hill, Careby, Stamford, Lincolnshire PE9 4EA, UK.

History of chemistry

Carl Scheele (1742–1786) and the discovery of oxygen

from E. L. Scott

THE birthday of oxygen is generally taken to be 1 August 1774, the day on which Joseph Priestley used a twelve-inch lens to focus the Sun's rays on the red calx of mercury (HgO), thereby liberating a gas in which "a candle burned...with a remarkably vigorous flame" (Priestley, J. in *Experiments and Observations on Different Kinds of Air* Vol. ii, 34; Johnson, London, 1775). But at least two years before that the Swedish chemist Carl Wilhelm Scheele, who died two hundred years ago, had already isolated oxygen. Moreover, Scheele's discovery was a confirmation of his belief in the existence of the gas, whereas Priestley's discovery was

totally unexpected.

Scheele's experiments and conjectures are described in his only book, *Chemische Abhandlung von der Luft und dem Feuer* (Uppsala and Leipzig, 1777), translated into English in 1780 (Scheele, C.W. *Chemical Observations and Experiments on Air and Fire* (trans. Forster, J.R.) Johnson, London, 1780). Within a few years the French chemist Lavoisier named the gas oxygène, believing it to be present in all acids, and had demonstrated its role in combustion. Previously, for more than a century, burning and the calcination of metals had been explained in terms of phlogiston — a 'principle' present in all metals and inflammable non-metals. On the application of sufficient heat, the phlogiston was thought to be attracted by the surrounding air, leaving a calx in the case of a metal and an acid when the substance burned was a non-metal.

Scheele, in several experiments carried out in closed vessels observed the diminution in volume of the air and the failure of the residual air to support combustion, processes now recognizable as oxidation. He concluded that air consists of two

Metal – phlogiston	→	calx
Non-metal – phlogiston	→	acid
Metal + oxygen	→	basic oxide
Non-metal + oxygen	→	acidic oxide

A comparison of the phlogiston and oxygen theories of combustion. Lavoisier, like his predecessors, did not distinguish between an acidic oxide and the acid which it forms on combining with water.

Ecology

Lunar cycles in lake plankton

from Geoffrey Fryer

RHYTHMIC behaviour is an almost universal attribute of animals, the most familiar example being that of activity by day and resting by night, or vice versa. A cycle related to the movements of heavenly bodies is also shown by many intertidal animals, the lunar cycle being associated with that of the tides. These rhythms are sometimes intrinsic (or endogenous): they continue, at least for a time, when the cycle is broken experimentally. More puzzling are rhythms associated with the lunar cycle shown by certain freshwater animals, of which only a few cases are known. Some African insects with aquatic larvae emerge only at certain phases of the Moon. Maciej Gliwicz (*Ecology* 67, 883; 1986) now demonstrates a lunar cycle in planktonic crustaceans, a cycle that is not intrinsic but is imposed on the organisms by a predator. The study reveals a fascinating set of ecological relationships; it remains to be seen whether as yet undiscovered adaptations are involved.

In the Cahora Bassa Reservoir, Zambezi Valley, Mozambique, four cladocerans (a ctenopod and three anomopods; see figure) and two copepods (a calanoid and a cyclopoid) show changes in numerical abundance that are clearly synchronized with the lunar cycle. Such a pattern could only be revealed by frequent sampling. Monthly sampling might greatly over- or underestimate numbers, depending on the starting date, and might fail to detect the very considerable fluctuations in population size.

How can this cycle be explained? Birth rates are such that the fluctuations cannot be attributed to changes in the fecundity of these rapidly reproducing organisms. Maximum death rates, however, occur around full Moon and changes in death

rates are clearly responsible for the fluctuations. The culprit is *Limnothrissa miodon*, a sardine-like clupeid fish native to Lake Tanganyika. Introduced to Lake Kariba, it has moved downstream and colonized Cahora Bassa. It feeds avidly on planktonic crustaceans, and calculations show that its depredations can account for even the most drastic declines observed in its planktonic prey.

There are changes in abundance of *Limnothrissa* adults, evidently short-lived in Cahora Bassa; annual changes in zooplankton abundance are related to these, but lunar fluctuations are not. Lunar fluctuations are caused by changes in feeding efficiency of the predator at different phases of the Moon. *Limnothrissa* feeds intensively throughout the night at full Moon, following the vertically migrating plankton to the surface: on moonless nights it is too dark for efficient hunting, so the predator disperses over a wide range of depths. *Diaphanosoma* and *Ceriodaphnia* begin to decline a few days before full Moon, as one might expect. In *Bosmina* and *Daphnia*, however, maximum death rates tend to occur after full Moon. Why should this be?

Gliwicz finds that stomachs of *Limnothrissa* are always fuller soon after full Moon than just before it, despite the similar light intensities and the fact that plankton is always more abundant before than after full Moon. Feeding intensity depends on the time of moonrise: before full Moon there is moonlight as soon as the sun sets and the plankton does not rise very near the surface; after full Moon there is complete darkness for between one and three hours, during which time the plankton ascends to the surface. When the full moon rises, its appearance is particularly sudden in Cahora Bassa, which lies in a gorge. Plankton that has previously risen to near the surface is thus suddenly exposed to predation by *Limnothrissa* and is heavily cropped. *Daphnia* and *Bosmina* are apparently particularly vulnerable to this trap. This is the first time that such lunar cycles have been demonstrated, but Gliwicz thinks he can detect others in published data on planktonic Cladocera in temperate lakes — cycles probably overlooked at the time as inherent sampling variations.

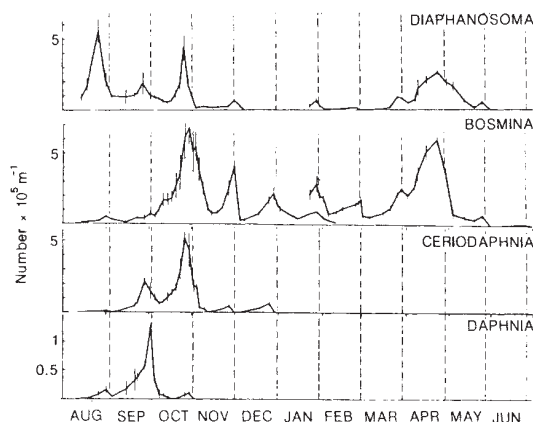
Since the classical work of Hrbáček (*Verh. int. Ver.*

Limnol. 13, 394; 1961) and of Brooks and Dodson (*Science* 150, 28; 1965) which shows how the introduction of a predator can alter the species composition of a plankton community, predation has been shown to have profound selective effects on the size, transparency and even on the gross morphology of planktonic anomopod cladocerans. The avoidance of predators has also long been considered to be a possible explanation of the phenomenon of vertical migration, and seems in some cases certainly to be involved. But dramatic fluctuations in abundance related to the lunar cycle suggest that planktonic crustaceans have not been able to evolve a complete response to increased predation at the time of full Moon. Avoidance of the surface waters on moonlit nights certainly reduces the effects of predation, but the trap set by the timing of sunset and moonrise is more difficult to detect and avoid.

Are such problems real or do they exist only in the human mind? Planktonic animals certainly suffer enormous mortality when caught in the Moon trap — yet they survive, and in sufficient numbers to build up their populations again within a lunar month. Fecundity may be the adaptive mechanism here: it might be advantageous to purchase grazing rights in rich pastures at the high price of periodic heavy payments.

The patterns of emergence of two African insects may be relevant to this hypothesis. In Lake Victoria the mayfly *Povilla adusta* emerges with a sharp peak of abundance on the second to fourth nights after full Moon (Hartland-Rowe, *R. Revue zool. Bot. afr.* 63, 185; 1958); in Lake Bangweulu the midge *Chironomus brevibucca* does so on the third to fifth nights (Fryer, *G. Bull. ent. Res.* 50, 1; 1959). Such synchronized emergence has the obvious advantage of facilitating the finding of mates. Swarming and spawning of marine organisms, essentially polychaetes, with similar lunar rhythms takes place at similar Moon ages. P. Koringa (*Mem. geol. Soc. Am.* 67, 917; 1957) has shown how marine animals might use moonlight to time and to synchronize their maturation, and this explanation fits the observed emergence times of *P. adusta* and *C. brevibucca*. In both species emergence takes place very near the phase of the Moon when the insects are most vulnerable to predation, and both are eaten extensively. Again, however, these species may find it advantageous to pay this price to achieve synchronized emergence at a convenient phase of the Moon. In the case of *C. brevibucca*, the numbers emerging are so immense that even heavy predation is trivial. □

Geoffrey Fryer is at the Windemere Laboratory of The Freshwater Biological Association, Ambleside, Cumbria LA22 0LP, UK.



Changes in population size of cladocerans in relation to the lunar cycle in Cahora Bassa Reservoir. Vertical broken lines, full Moon.

Cognitive neurophysiology

Signs of language in the brain

from John C. Marshall

THOSE of us with intact hearing are often inclined to dismiss the manual communication of deaf people, discussed in a study reported on page 363 of this issue¹, as a mishmash of pantomime and iconic signals, eked out by finger spelling. In many teaching institutions committed to strict 'oralism', signing was (and in some instances still is) firmly suppressed, with disastrous consequences for the deaf child's educational attainment²; even in academic discussion, sign language was frequently regarded as, at best, a primitive pidgin, and not a 'real' language.

Yet in Europe, true sign languages, with all the formal complexity and expressive power of spoken language, are known to exist in (at least) Austria, Denmark, Finland, France, Germany, Norway, Portugal, Rumania, Sweden, Switzerland, the Soviet Union, the United Kingdom and Yugoslavia. These different, often mutually unintelligible languages, which have little or no superficial similarity with the local spoken languages, have been passed down through the generations from deaf parents to their deaf (and often hearing) children.

American Sign Language (ASL) branched off from French Sign Language some two hundred years ago, and the current intense interest in the linguistic structure of ASL derives in large measure from the pioneering work recently undertaken by Ursula Bellugi and Edward Klima³ at the Salk Institute in California. Once the formal structure of manual signing had been elucidated — how hand configuration,

position and movement through space (including change in hand configuration) combine to express vocabulary and syntax — it was clear that ASL is indeed a fully-fledged language that can serve every day conversation, intellectual argumentation, wit and poetry⁴. But now there arises an interesting issue about the neuronal substrate for sign.

The basic complementary organization of the duplex human brain is between speech processing (committed to the left

sphere damage, signing is little disturbed, although many non-language visuo-spatial tasks (including drawing) may show severe impairment^{5,6}. It would seem, then, that it is language *per se* (not solely speech perception and production) that is lateralized to the left cerebral hemisphere. Grammatical structure is processed by the left hemisphere even when the overt physical realization of that structure involves a visuo-spatial medium.

The case now reported by Damasio, Bellugi, Damasio, Poizner and Van Gilder⁷ demonstrates in even stronger fashion that the above conclusion is warranted. Their patient, a young woman with normal hearing, spoke English but was also proficient in ASL as a second language; she earns her living as an inter-

preter and counsellor for deaf people whose native language is ASL. Unfortunately, she suffered increasingly from frequent partial complex seizures that could not be controlled pharmacologically. Surgical intervention — resection of the temporal lobe — was therefore considered. But which temporal lobe (left or right) should be ablated? Bilateral temporal lobectomy results in a gross amnesic syndrome that precludes any kind of normal life; removal of the language-dominant temporal lobe could result in aphasic impairment that would destroy the patient's livelihood.

The application of advanced *in vivo* brain imaging techniques showed anatomic and metabolic

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Movement trajectories of grammatical processes (courtesy of U. Bellugi).

cerebral hemisphere) and visuo-spatial together with visuo-manipulative processing (committed in the main to the right hemisphere). A central question in the psychobiology of language is thus whether the underlying specialization of the left hemisphere is restricted to auditory-vocal language (that is, speech), or rather includes a more 'abstract' representation of linguistic form (grammar), irrespective of the mode in which it is expressed⁸.

Natural sign languages provide a privileged insight into this question. Such a system simultaneously qualifies as a language and as an extremely complex and precise set of gestures, executed in space and perceived visually. Nonetheless, studies of disorders in the expression and comprehension of sign consequent upon unilateral brain damage show that, just as with aphasia for spoken language, it is left hemisphere damage that provokes these sign language aphasias^{6,7}; by contrast, in deaf patients with extensive right hemi-

asymmetries consistent with normal left hemisphere processing of speech in the patient. More direct evidence for the control of language skills was then obtained by barbiturate injection. The brief inactivation of frontal and central areas of one hemisphere (which include language-committed cortex) can be achieved by injection of sodium amytal into the ipsilateral carotid artery. When this was performed, a reversible lesion (pharmacologically induced) of the left hemisphere produced in the patient a temporary aphasia for English and ASL; left cerebral dominance for both language systems was thereby indicated, and the severity and the duration of the sign language impairment was even greater than in spoken English. Subsequent neurosurgical intervention — partial resection of the right temporal lobe — was found, three and twelve months post-operatively, to have produced no neuropsychological impairment of either English or sign language.

1. Damasio, A., Bellugi, U., Damasio, H., Poizner, H. & Van Gilder, J. *Nature* **322**, 363 (1986).
2. Conrad, R. *The Deaf School Child: Language and Cognitive Function* (Harper and Row, London, 1979).
3. Klima, E.S. & Bellugi, U. *The Signs of Language* (Harvard University Press, Cambridge, 1979).
4. Klima, E.S. & Bellugi, U. *Sign Language Studies* **8**, 204, (1975).
5. Bellugi, U., Poizner, H. & Zurif, E. in *Neural Models of Language Processes* (eds Arbib, M.A., Caplan, D. & Marshall, J.C.) 217 (Academic, New York, 1982).
6. Kimura, D. *Sign Language Studies* **33**, 291 (1981).
7. Poizner, H., Bellugi, U. & Iragui, V. *Am. J. Physiol.* **246**, R868, (1984).
8. Kimura, D., Davidson, W. & McCormick, C.W. *Brain and Language* **17**, 359 (1982).
9. Poizner, H., Kaplan, E., Bellugi, U. & Padden, C. *Brain and Cognition* **3**, 281 (1984).
10. Marshall, J.C. *Cognition* **17**, 209 (1984).

The patient continued to manipulate the spatial syntax of ASL with the same fluency she had shown pre-operatively. Dr A. Damasio reports (personal communication) that the patient has been seizure-free since the operation and has successfully returned to her demanding work as a sign-language interpreter.

These results, then, provide further testimony for the modular organization of the human brain; different domain-specific cognitive representations are computed by the two halves of the adult brain, irrespective of the physical form whereby

those representations are made manifest¹⁰. It is, however, possible that by virtue of their combined linguistic and visuo-spatial nature, the facile acquisition of signs requires both hemispheres to be fully intact. Damasio, Bellugi and their colleagues now intend to investigate how easily their patient can learn new signs, including the quite distinct space structures of Chinese Sign Language. □

John C. Marshall is in the Neuropsychology Unit, Neuroscience Group, Radcliffe Infirmary, Oxford OX2 6HE, UK.

Solid-state physics

New semiconducting polymers

from R.H. Friend

THE discovery in 1977¹ that the addition of small amounts of strong oxidizing or reducing chemicals to polyacetylene (doping) produces levels of conductivity typical of metals contradicted long-held views that electrons in organic semiconductors necessarily have low mobilities. Progress since then has been impeded by the limited availability of polymers with desirable properties and also by difficulties in processing them. This constraint has been loosened by S.A. Jenekhe, who on page 345 of this issue² reports the smallest energy gap yet obtained in an organic polymer, which brings closer the use of this class of compounds as semiconductors and transparent conductors.

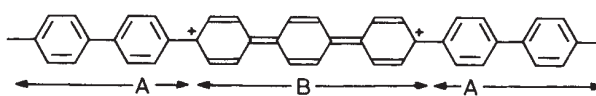
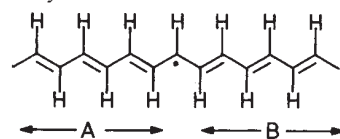
The feature common to the organic semiconductor class of polymers is the presence of p_z orbitals on each carbon atom on the chain, giving a conjugated structure which can be represented (slightly misleadingly) by alternation of 'double' and 'single' bonds. Overlap of adjacent p_z orbitals gives well-delocalized π electron wavefunctions and high electronic mobilities. These polymers do not conform to the models that work for inorganic semiconductors.

Whereas it can be said that three-dimensionally bonded inorganic semiconductors are semiconductors by accident (that is, there happens to be an energy gap between the highest occupied and lowest unoccupied levels), the same is not true for 'one-dimensional' polymeric semiconductors, where it is energetically favourable to set up the interatomic spacings along the chain so that there is an energy gap between the highest π and the lowest π^* levels.

The semiconductor gap in the *trans*-isomer of polyacetylene is caused by alternation of bond lengths along the chain;

this dimerization introduces an energy gap between the occupied π 'bonding' states (which have high electron density in the short, double bonds) and the unoccupied π^* states (with electron density in the long, single bonds). The strength of the bond alternation, and hence the size of the energy gap, is determined by the trade-off between the cost in elastic strain energy associated with the distortion and the reduction in π electron energy that it achieves. The observed gap is approximately 1.5 eV.

The particular novelty in *trans*-polyacetylene is that the sense of bond alter-



Bond-alternation defect (shown as the neutral defect) separating the degenerate A and B phases of *trans*-polyacetylene (above); and a positively charged bipolaron defect on a poly(phenylene) chain, showing the quinonoid character within the defect (B phase) and benzenoid character (A phase) elsewhere on the chain (below).

nation is not determined, so that there are two, degenerate phases, labelled A and B in the figure. In consequence the chain can support stable defects in the bond alternation sequence, separating the A and B phases as shown in the figure. These topological defects, which have an associated 'non-bonding' p_z level, are calculated to move along the chains as low-mass particles and possess most of the qualities of solitary waves (solitons)³. Furthermore, the result of adding charge to the chain, through chemical doping or following photoexcitation, is always to create these defects, which can be positively charged

(p_z level empty, spin 0), neutral (one electron on p_z level, spin $1/2$) or negatively charged (p_z level doubly occupied, spin 0). This is in contrast to the doping of traditional, inorganic semiconductors in which added charges are present as holes in the valence band or electrons in the conduction band.

For most of the other conjugated polymers that have been synthesized, the sense of bond alternation along the chain is determined, and the A and B phases are non-degenerate. For the polyphenylene chain illustrated in the figure, the benzenoid (A) phase is lower in energy than the quinonoid (B) phase. The usual consequence of this is that there is now an additional contribution to the semiconductor gap (π levels are of benzenoid character, π^* levels quinonoid), and indeed this is true in many cases, including polyphenylene (3 eV) and polythiophene (2 eV). Soliton defects are no longer stable, although in pairs they can form 'polaron' defects with a short length of B phase on the chain. Although the polaron defect is not stable on a neutral chain, it does provide the cheapest way of accommodating charges added to the chain, either as a singly charged polaron or as a doubly charged bipolaron⁴ (see figure).

Hence the discovery of conjugated polymers with band gaps smaller than that of polyacetylene was unexpected. To obtain these small band gaps it is necessary to design the polymer so that the energy difference between the A and B phases is as small as possible, and this is apparently achieved in polyisothionaphthene⁵ which has a gap of slightly more than 1 eV. An alternative strategy is to design a polymer which is forced to have alternating sequences of the A and B phases along the chain so that there is little difference in the benzenoid/quinonoid character in the π and π^* states. Jenekhe has achieved this for a class of thiophene-derived polymers that reduce the gap to as low as 0.75 eV.

Narrow-gap conjugated polymers will extend the scope both for applications and for fundamental investigations. Much as for inorganic semi-conductors, narrow-gap materials should show high electronic mobilities. In the context of the polymers, the soliton and polaron defect states will be delocalized and have low effective masses. This is desirable for many semiconductor applications and, in addition, these materials can be transparent in the visible part of the spectrum and are thus of use as transparent conductors⁶.

Beside the tuning of electronic properties that is now achieved with new synthetic routes, processibility via a soluble phase, either a 'precursor' polymer or a conjugated polymer with soluble side

groups⁶, is now routine. Well-crystallized and highly oriented films of several precursor-route polymers can be obtained by stretch orientation during transformation; these films have allowed the first experimental investigations of the anisotropy of electron motion along or between chains^{7,8}. Although electron motion along chains is easy, it is important to recognize that the degree of contact between the π electron wavefunctions on adjacent chains is quite strong, and that interchain electron motion is not difficult. Experiments using oriented polyacetylene⁸ demonstrate that the quantum yield for photo-generation of charged pairs is higher when the band-gap illumination is polarized perpendicular, rather than parallel, to the

chains, so that the photons absorbed transfer electrons between chains. The scope for development of the highly dichroic properties of these materials is vast. □

1. Chiang, C.K. *et al.* *Phys. Rev. Lett.* **39**, 1098 (1977).
2. Jenekhe, S.A. *Nature* **322**, 345 (1986).
3. Su, W.P., Schrieffer, J.R. & Heeger, A.J. *Phys. Rev. B* **22**, 2099 (1980).
4. Campbell, D.K., Bishop, A.R. & Rice, M.J. in *Handbook of Conducting Polymers* (ed Skotheim, T.J.) (Dekker, New York, 1986).
5. Wudl, F., Kobayashi, M., Colaneri, N., Boysel, M. & Heeger, A.J. *Molec. Cryst. Liq. Cryst.* **118**, 199 (1985).
6. Calvert, P. *Nature News and Views* **304**, 487 (1986).
7. Kahlert, H. & Leising, G. *Synth. Metals* (in the press).
8. Townsend, P.D. & Friend, R.H. in *Synth. Metals* (in the press).

R.H. Friend is in the Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK.

Meteorites

Window on the early Solar System

from Derek Sears

It is an oft-repeated truism that meteorites are a window on the early Solar System. They are, however, complex: the astronomer B. Pagel once remarked, after taking a cursory glance through the window, that "the oracle is . . . wrapped in obscurity of truly Delphic proportions"¹. Fortunately, Pagel was unduly pessimistic, because the window sometimes clears to present a unique and fascinating picture of places and processes as they were 4,600 million years ago, as shown in a recent paper by David Wark².

One of the most important developments in the study of meteorites over the past few decades is the realization that some meteorites retain a better record of the nebular phase of their history than others; these primitive or type 3 chondrites contain many individual components with particular significance for deciphering the earliest history of solid matter in the Solar System. Some of the better-studied components are the silicate aggregates called chondrules³; the fine-grained, almost smoke-like, matrix⁴; certain kinds of metal grain⁵; and the graphite-magnetite aggregates⁶. But the objects which have attracted the most attention are calcium-aluminium-rich inclusions (CAIs)⁷, one of which is the subject of Wark's paper.

Wark describes a CAI discovered in the Allende meteorite shortly after it fell in 1969. The inclusion (CAI 3643) consists of a core, mantle and crust. The crust appears to have been formed by rapid episodic condensation and aggregation in a compositionally unchanging gas. The core consists predominantly of hibonite, a mineral which is one of the first to condense in a gas of solar composition, whereas the mantle and crust consist predominantly of melilite, which is thought to form by re-

action between the hibonite and the gas. The core has the texture of a material which formed by deposition from a vapour. There are many metal nuggets in the centre of the core which consist of highly refractory trace elements, whereas those in the outer parts of the core contain elements of less refractory nature — it is as if they mopped up the elements missing from the nuggets at the core's centre. Other refractory trace elements, the rare-earth elements, show a steady decrease in abundance with increasing volatility, as if the formation of the CAI occurred over the same temperature interval as the elements were condensing.

All these details are consistent with an episodic condensation and accretion of the core, with the mantle and crust being made by later episodes of condensation onto the earlier-formed solids. The processes occurred at high temperatures and involve solids which are refractory to various degrees. Once formed, the entire assemblage suffered several alteration processes at much lower temperatures, with evidence for metamorphism, metasomatism and the introduction of a vapour rich in halogens, alkali metals and iron.

Many who have worked on this CAI have suggested that the inclusions, or some material in them, were not formed in our Solar System, but were transported here from interstellar space during or just before the formation of the Solar System. This idea is rooted in the unusual isotope ratios frequently encountered in the inclusions. It was long believed that the isotopes in our Solar System became thoroughly mixed during its formation. CAIs also contained several short-lived radioactive isotopes at the time they were formed. Some of the nuclides have very short half-lives (for example ²⁶Al, 0.72

million years), which has suggested⁸ that the event which ejected the isotopes into the interstellar medium (say, a supernova explosion) was that which triggered the formation of the Solar System. Thus, the condensation sequences would have occurred in the expanding and cooling shells of ejected supernova material⁹.

An alternative view is that interstellar material, with essentially normal chondritic composition, entered the primordial solar nebula at cosmic velocities and became heated in a manner analogous to meteorites entering the atmosphere¹⁰. The evaporation sequence as the material is heated would be the reverse of the condensation sequence, so that CAIs can be regarded as distillation residues rather than high-temperature condensates¹¹.

Wark points out that the properties of CAI 3643 are not consistent with ideas that involve exposure to various chemical environments. Although the zones in the inclusion, and the radial gradients in mineralogy and refractory-element chemistry, imply different temperature regimes for the aggregation processes, they must have occurred quickly for the differences to have persisted, and they must have occurred in a gas which did not change very much in its composition. Instead, Wark suggests that the CAI formed in a single place and over a fairly restricted time interval; he favours the formation of the CAI by episodic condensation and aggregation during localized transient heating events in the asteroid belt.

Wark's paper will not lay to rest the idea that CAIs are extra-Solar System in origin. It is always possible, although unlikely, that the various extra-Solar System atmospheres that have been postulated were compositionally very similar and movement from one to another was very rapid. It is also possible that not all CAIs were created in the same way. But Wark provides a clear indication that, in at least this instance, the condensation and accretion sequence was followed episodically and in a very restricted regime of time and environment. In this sense, our window on the early Solar System has cleared a little, and provided us with a view that can be obtained in no other way. □

1. Pagel, B. *Observatory* **99**, 162 (1979).
2. Wark, D.A. *Earth planet. Sci. Lett.* **77**, 129 (1986).
3. King, E.A. (ed.) *Chondrules and Their Origins* (Lunar and Planetary Institute, Houston, 1984).
4. Huss, G.R., Keil, K. & Taylor, G.J. *Geochim. cosmochim. Acta* **45**, 33 (1981).
5. Rambaldi, E.R., Sears, D.W. & Wasson, J.T. *Nature* **287**, 817 (1980).
6. Scott, E.R.D., Taylor, G.J., Rubin, A.E., Okada, A. & Keil, K. *Nature* **291**, 544 (1981).
7. Grossman, L. *A. Rev. Earth planet. Sci.* **8**, 559 (1980).
8. Cameron, A.G.W. & Truran, J.W. *Icarus* **30**, 447 (1977).
9. Lattimer, J.M., Schramm, D.N. & Grossman, L. *Astrophys. J.* **219**, 230 (1978).
10. Wood, J.A. *Earth planet. Sci. Lett.* **70**, 11 (1981).
11. Wood, J.A. *Earth planet. Sci. Lett.* **56**, 32 (1981).

Derek Sears is in the Department of Chemistry at the University of Arkansas College of Arts and Sciences, Fayetteville, Arkansas 72701, USA.

Geometry

Exotic structures on four-space

from Ian Stewart

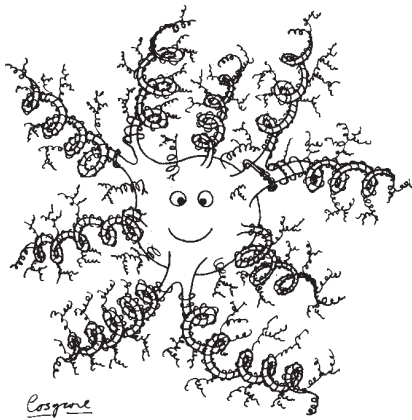
ONE of the biggest surprises in topology was the discovery in 1982 that four-dimensional euclidean space does not have a unique structure as a differentiable manifold. The existence of an 'exotic' structure on four-space was proved by Michael Freedman, by combining his techniques for analysing four-manifolds (*J. diff. Geom.* **17**, 357; 1982) with a new and startling result due to Simon Donaldson (*J. diff. Geom.* **18**, 279; 1983). A question that immediately arose was how to classify all such structures: in particular, how many are there? Robert Gompf (*J. diff. Geom.* **18**, 317; 1983) found four distinct differentiable structures — the standard one plus three exotics. A fifth 'universal' structure (containing all the others) was found by Freedman and Larry Taylor. And now Gompf has shown (*J. diff. Geom.* **21**, 283; 1985) that there exists an uncountable infinity of exotic structures on four-space.

A differentiable manifold is a multi-dimensional analogue of a smooth two-dimensional surface. The idea is of central importance in mathematics and has applications throughout the physical sciences. For example, the motion of the Moon, the Earth and the Sun under their mutual gravitational influence corresponds to the motion of a point on an 18-dimensional 'phase space' having three dimensions of position and three of momentum for each body. The original idea goes back to Bernhard Riemann in 1854; in its modern form it is due to Hermann Weyl in the 1930s. An n -dimensional manifold is a collection of overlapping 'patches', each looking like a piece of n -dimensional euclidean space. Each patch inherits a coordinate system, which means that on each overlap there are two distinct coordinate systems — one from each patch. These do not necessarily agree; but if each is related to the other by a differentiable transformation, then the system of patches is said to define a differentiable structure on the manifold.

For example consider an ordinary two-dimensional sphere. This can be considered as two overlapping patches, each forming rather more than a hemisphere. For example, in geographical terms, let one patch be everything south of the Tropic of Cancer, and let the other be everything north of the Tropic of Capricorn. The overlap is an equatorial band between the two tropics. Each patch is a distorted circular disk, and it is not hard to find coordinate systems on the disks that provide smooth overlaps. An n -dimensional hypersphere can be defined by an analo-

gous construction, using two balls in n -dimensional space.

It is easy to see that a given manifold can have many different differentiable structures, that is, systems of smoothly overlapping patches. But the obvious examples are related in an essentially trivial way. Start with a manifold M and a differentiable structure S . Take any continuous one-to-one correspondence of M with itself and apply it to all of the patches describing S , to get a new set of patches S' . Then, in general, S' is different from S , but the patches of S' fit together to give the same manifold as those of S . To exclude trivial changes of this type, say that



Artist's impression of an exotic four-space. The structure resembles ordinary four-space at the centre, but becomes more and more complicated as it approaches infinity, putting out branched tentacles in all directions. This conjectural picture compresses four dimensions into two, and is correct, if at all, only in spirit. An explicit description of the true geometry is not yet known.

under these circumstances S and S' are 'equivalent'. Then a more interesting question arises: can there be several inequivalent differentiable structures on the same manifold?

Early results suggested not: for example, it was proved in the 1950s that if the manifold is three-dimensional then there is a unique differentiable structure up to equivalence. However, matters took an unexpected turn in 1963 when John Milnor proved that the seven-dimensional hypersphere has at least two inequivalent differentiable structures: the 'standard' one and another exotic one. A detailed study of these exotic spheres revealed that there are exactly 28 inequivalent differentiable structures on the 7-sphere.

Even so, it was not expected that this kind of trouble would show up in ordinary euclidean n -space. Indeed it was known

that for $n \neq 4$ there is a unique differentiable structure. The conjecture that uniqueness holds for all euclidean spaces lay at the heart of smoothing theory, a method for studying non-differentiable manifolds by providing them with a differentiable structure. So it was quite a shock when Freedman and Donaldson showed that four-dimensional space behaves in a completely different and rather horrifying way. There exist 'exotic R 's', that is, differentiable structures on four-space that are not equivalent to the standard one (see figure).

The problem is to tame these unruly structures by imposing some comprehensible order on them. Which brings us back to the question: how many exotic R 's are there? Mathematicians distinguish many different 'sizes' of infinity. The cardinal of a set is the size of infinity that corresponds to it. The smallest cardinal, called aleph-zero, or countable infinity, is that corresponding to the set $\{1, 2, 3, \dots\}$ of positive integers. A set is countably infinite if it can be put into one-to-one correspondence with the positive integers. Many sets — pairs of integers, the rational numbers, the algebraic numbers — are countably infinite. But the set of real numbers cannot be put into one-to-one correspondence with the positive integers, and is said to be uncountably infinite. There is a consistent arithmetic of infinite cardinals: it has its own bizarre logic, and there are many pitfalls. For example, it is possible for one set to contain another (and thus be 'bigger') even though both sets have the same cardinal.

Gompf first found a family $R_{s,t}$ of exotic R 's, where s and t are positive integers, with the property that $R_{s,t}$ embeds in $R_{u,v}$ if and only if $s \leq u$ and $t \leq v$. In particular $R_{s,t}$ is equivalent to $R_{u,v}$ only when $s = u$ and $t = v$. This provides a countably infinite set of inequivalent exotic R 's. Clifford Taubes has now developed a generalization of Donaldson's results, inspired by an observation of Freedman that such a generalization would imply the existence of an uncountable infinity of exotic R 's. This idea can be applied to Gompf's construction, and it yields an even more extensive family of exotic R 's. It is probable — but not yet certain — that these results give the correct cardinal for the set of all exotic R 's. That does not mean that Gompf's examples exhaust all possibilities, because 'larger' sets can have the same cardinal. Indeed Taubes's set is smaller than Gompf's, yet the cardinal is the same in both cases.

The set of all exotic R 's also carries a natural algebraic operation, the 'end-sum'. There are a number of analogous operations in topology. The simplest combines two knots to yield their 'sum' by tying both in the same piece of string. Similarly the connected sum of two manifolds can be obtained by cutting a small

disk out of each and gluing a tube between the holes that result. (Milnor showed that such a construction applied to his 28 exotic 7-spheres turns them into a very nice algebraic object: a group.) The end-sum of two exotic R's is obtained by 'gluing them together at infinity'. To find an abstract description of the structure of the set of all exotic R's under the end-sum operation is a problem of central importance. It would not only classify them all, but also explain how they relate to each other.

The importance of these ideas for mathematics is manifest. We know that four-dimensional space differs in a dramatic and rich way from all other dimensions. It is obviously of fundamental interest to explore exactly how and why this occurs.

The relevance to applied science is not

so clear: the 'standard' structure on four-space is after all the one suggested by physical observation. But the influence of mathematics and applications on each other, while profound, is seldom immediate and hardly ever direct. For example the existence of exotic R's is established by combining some recent ideas in quantum physics — gauge field theories — with some extremely abstract techniques of modern topology. The same combination is appearing in basic questions about unifying the forces of nature. A fundamental mathematical truth is important because, once proved true, it stays true. Exotic R's are abroad in the world, and the world is changed — forever. □

Ian Stewart is at the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

Computational cosmology

Galactic haloes need computers

from George Efsthathiou

OBSERVATIONS of rotation in spiral galaxies suggest the presence of dark material in haloes surrounding the luminous component. Little is known about the nature of this dark matter — it could be black holes, ultra-small 'stars' or even elementary particles such as massive neutrinos, photinos or gravitinos. However, the way that the rotation speed varies with distance from the centre of a spiral suggests that the mass in the halo increases linearly with radius. This behaviour is important for at least two reasons. The first concerns the origin of the observed spatial distribution and whether it arises naturally from small perturbations which grow by the action of gravity. The second concerns the total extent of the dark haloes, and whether there is enough dark matter so that the Universe will eventually collapse back to a singularity. The first of these questions is addressed by P.J. Quinn, J.K. Salmon and W.H. Zurek elsewhere in this issue (*Nature* 322, 329; 1986) and both questions were among those discussed at a recent meeting*.

Quinn *et al.* use computer simulations which model the gravitational interactions between a large number of mass points to study the formation of structure in an expanding universe under various assumptions concerning the initial conditions. If the mass points are initially distributed at random (white noise), then the haloes that form in the simulations are more centrally concentrated than indicated by the observations. But the authors also show that density fluctuations with significant contributions from long-wavelength perturbations (more like 'pink noise'), develop

haloes with a linear mass distribution resembling that around spirals.

Interestingly, the pink-noise irregularities arise naturally if the dark matter is in the form of weakly interacting massive particles such as gravitinos. The evolution of structure in such models has been investigated using computer simulations by C.S. Frenk *et al.* (*Nature* 317, 595; 1985) who showed that the model broadly accounts for the distribution of mass in haloes and their abundance in space.

But what relation do these studies have to the models of galactic distribution and clustering that follow from cosmic string theories? There are very significant differences: the underlying assumption of the gravitating mass-point studies is that statistics of the initial fluctuations are gaussian, whereas strings introduce considerable skewness — which may lead to clustering and superclustering; furthermore numerical models of the strings developed so far exclude gravitational forces. The two types of study clearly need to be combined in a single model, but the computer code to do the job has not yet been written — largely because of the power and time that are necessary to handle simultaneously gravitating mass points, strings and internal string forces.

Why do we need computer simulations at all in such problems? The answer lies in the non-linear nature of gravitation. In general, it is not possible to study the evolution of a gravitating system of mass points by analytic methods unless one adopts highly restrictive assumptions such as spherical symmetry. Furthermore, in the cosmological context described above, halos do not evolve as isolated systems. Rather they form through a complicated

series of mergers between smaller sub-units. It seems unlikely that we will readily develop an analytic theory capable of reliably modelling such events. Yet can we believe computer simulations which typically contain less than 50,000 mass points, some six orders of magnitude less than the number of stars in one galaxy? Does the artificial 'graininess' or the presence of boundaries in computer simulations adversely affect the results? The answer is yes in some cases and no in others, and it is up to the numerical experimenter to convince his readers of the reliability of key conclusions by applying appropriate tests. Sadly, this has not always been done in the past and many researchers, in other fields as well as in cosmology, remain sceptical of the computer simulations.

There are also other issues. In the present mass-point studies, essentially any 'clump' that forms is considered to be a galactic halo. The familiar flattening of galaxies into disks does not occur because it requires dissipating forces, such as radiative cooling from high-density gas. Consideration of these effects requires a hybrid program that can handle both gas dynamics and gravitational mass-point interaction. Since a full model must also take into account feedback effects where star formation can raise intergalactic gas temperatures from microwave background levels to 10,000 K, the ultimate computer model of galaxy formation is bound to be highly complex.

Much larger simulations could be achieved by using improved algorithms, a new generation of supercomputers, or both, as discussed at the meeting. Most present-day computers are simple uniprocessor machines: the current generation of supercomputers such as the Cray-XMP or Cyber-205 achieve a high performance by applying the same instruction simultaneously to many pieces of data which belong to a 'vector'. The US National Science Foundation has set up several centres for large-scale scientific computing where the computers will be coarse-grained parallel machines such as the ETA-10. This enormous computer consists of 8

100 years ago

Animal Intelligence

A REMARKABLE instance of animal intelligence has lately come under my notice. In a neighbour's bungalow in this district two of our common house-swallows (*Hirundo javanica*) built their nest on top of a lamp that hangs in the dining-room. As the lamp is either raised or depressed by chains fixed to a central counterweight, these chains pass over pulleys fixed to a metal disk above, on which the nest was placed. The swallows built over each pulley a little dome, allowing sufficient space, both for wheel and chain to pass in the hollow so constructed, without danger to the nest, which was not only fully constructed, but the young birds were reared without further danger.

From *Nature* 34, 265; 22 July 1886.

*Use of computers in stellar dynamics, 2-4 June 1986, Institute for Advanced Study, Princeton, New Jersey, USA.

Cyber-205 machines with a massive shared memory. At the other end of the spectrum, fine-grained parallel machines such as the Connection Machine use many simple processors which communicate with each other via a complicated network of interconnections.

The optimal algorithm for solving the gravitational n -body problem clearly depends on the architecture of the machine. Existing programs such as Aarseth's famous n -body code can easily be adapted for vector machines and will be used extensively at the supercomputer centres. Algorithms for very fine-grained computers (which use many simple processors) are likely to be quite different. Several ingenious possibilities were described at the meeting (Joshua Barnes, Institute for Advanced Study, Princeton).

Very few cosmologists and astrophysicists have thought seriously about building their own special-purpose supercomputer, yet it is feasible to acquire the

necessary skills (Gerald Sussman, MIT). Sussman also argued that it is now possible to build one-off machines to study the stellar dynamical evolution of, say, an entire globular star cluster containing one million stars.

But even the globular cluster problem is not entirely stellar dynamical. At some point during evolution, the finite sizes of stars may become important, as may mass-loss and other astrophysical effects. In the case of cosmological simulations, our lack of knowledge of gas dynamics and star formation impose severe limitations on our understanding. Supercomputers will undoubtedly make a major impact in cosmology and astrophysics, but in many cases they will solve only idealized problems. Theorists will still have plenty to do in applying the results to real stellar systems. □

George Efsthathiou is at the Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK.

Nitrogen fixation

A role for vanadium at last

from R. Cammack

FIFTY years after Bortels reported¹ that vanadium could substitute for molybdenum as a trace element essential for nitrogen fixation in the soil bacterium *Azotobacter vinelandii*, Robson *et al.*, on page 388 of this issue², report conclusive evidence for an alternative vanadium-containing nitrogenase from *A. chroococcum*, a result made possible by recent developments in the genetics of nitrogen-fixing bacteria. Robson *et al.* produced a mutant strain in which the structural genes for the normal molybdenum-containing nitrogenase are deleted and demonstrate that the alternative nitrogenase uniquely requires vanadium for its synthesis, and in stoichiometric amounts.

The complex, nitrogen-fixing enzyme nitrogenase has two types of protein component, both containing iron-sulphur clusters. One of these components, the iron-molybdenum protein, also contains a cluster of unknown structure, the iron-molybdenum cofactor (Fe-Mo-Co): there is compelling evidence that this is the site at which dinitrogen is reduced to ammonia. The function of the other component is to 'pump' electrons, using energy from ATP, into the iron-molybdenum protein. The vanadium-containing nitrogenase appears to have a very similar arrangement, a structure containing vanadium presumably substituting for the Fe-Mo-Co, although the proteins involved are the products of different genes.

Chemically, the substitution of vanadium for molybdenum makes good sense. The two elements are capable of

similar oxidation-reduction reactions at low redox potentials. Model compounds for the nitrogenase reaction that can catalyse the reduction of dinitrogen to ammonia have been synthesized, most containing molybdenum, although complexes of vanadium are also effective³.

Ironically, although nitrogenase was partially purified from vanadium-grown *A. vinelandii* about 15 years ago and several of its catalytic properties, such as decreased sensitivity to cobalt inhibition, correctly described, the significance of vanadium in this system was not recognized. This was partly because of the extreme difficulty of obtaining growth media that were sufficiently molybdenum-free, which meant that all nitrogenase preparations contained a certain amount of molybdenum. It must have seemed unreasonable at the time to assume that there existed a second nitrogenase system with almost identical protein structure to the molybdenum nitrogenase system, which rapidly decomposes during purification and has very little activity in the conventional acetylene-reduction assay. Instead it was concluded that vanadium had a 'sparing' effect, stabilizing the molybdenum nitrogenase when molybdenum was in short supply. The idea that all nitrogenases contain molybdenum became firmly established and molybdenum was routinely added to all growth media for nitrogen-fixing organisms. Because the synthesis of the vanadium nitrogenase is repressed even by traces of molybdenum, its presence was not detected. A recent

survey of the literature showed that Bortels' 1930 paper describing a requirement for molybdenum in nitrogen fixation has been cited 55 times, while his 1936 paper⁴ describing the alternative use of vanadium was cited 4 times.

The more recent discoveries stem from the work of Bishop⁵ who, in the face of much initial scepticism, demonstrated the presence of a second nitrogenase system in *A. vinelandii* grown in molybdenum-deficient conditions. He obtained a *nif* mutant which could not express the conventional molybdenum nitrogenase, but which could still grow by fixing nitrogen in the absence of molybdenum. He showed that these 'pseudo-revertants' were tungstate resistant and expressed a different nitrogenase whose activity (as measured by hydrogen production) was less sensitive to inhibition by acetylene, carbon monoxide or tungstate. It appears that it was the vanadium present as a contaminant in tungstate that allowed Robson *et al.*² to discover the vanadium enzyme.

Azotobacter synthesizes the molybdenum nitrogenase in preference to the vanadium nitrogenase when molybdenum is available presumably because the molybdenum enzyme is more stable, and possibly catalyses less of the wasteful side-reaction in which hydrogen is produced. Although molybdenum is difficult to remove from laboratory media, it may become a limiting factor in some soils, and hence the capacity to make the alternative nitrogenase confers a selective advantage. A question of great interest is whether other nitrogen-fixing organisms, and in particular the economically important rhizobia, are capable of making the vanadium nitrogenase.

Macara⁶ has described vanadium as "an element in search of a [biochemical] role". It is known that vanadium in extremely small amounts is a nutritional requirement for many types of organism, possibly including higher animals⁷. The sea-squirt *Ascidia nigra* contains high concentrations of V(III) ions in specialized blood cells. Vanadium is also accumulated by the toadstool *Amanita muscaria*⁸ and a vanadium-containing bromoperoxidase has recently been isolated from a seaweed, *Ascophyllum nodosum*⁹. With the establishment of vanadium in nitrogenase, its role is beginning to emerge. □

1. Bortels, H. *Zentbl. Bakt. Parasitkde Abt. II* **95**, 193 (1936).

2. Robson, R.L. *et al. Nature* **322**, 388 (1986).

3. Nikonova, L.A., Isaeva, S.A., Pershikova, N.I. & Shilov, A.E. *J. molec. Catal.* **1**, 367 (1975).

4. Bortels H. *Arch. Mikrobiol.* **1**, 333 (1930).

5. Bishop, P. *Trends biochem. Sci.* **11**, 225 (1986).

6. Macara, I.G. *Trends biochem. Sci.* **5**, 92 (1980).

7. Nechay, B.R. *Fedn Proc.* **45**, 123 (1986).

8. Gillard, R.D. & Lancashire, R.J. *Phytochemistry* **23**, 179 (1983).

9. De Boer, E., Van Kooyk, Y., Tromp, M.G.M., Plat, H. & Wever, R. *Biochim. biophys. Acta* **869**, 48 (1986).

R. Cammack is a Reader in the Department of Biochemistry, King's College London, Campden Hill Road, London W8 7AH, UK.

Observation of ^{110m}Ag in Chernobyl fallout

SIR—The initial radiometric flurry of interest in fission products with relatively short half lives following the Chernobyl reactor accident has been replaced with concern about the entry of fission products with longer half-lives into the food chain.

The movement and slaughter of lambs in Wales, Cumbria and Scotland has recently been prohibited due to the identification of lamb meat containing abnormally high levels of ^{134}Cs ($T_{1/2}=2$ yr) and ^{137}Cs ($T_{1/2}=30$ yr).

The radiometric laboratory of the Department of Physics at the University of Liverpool, which houses several low-level γ -ray counting systems, has been employed to carry out measurements on various dairy and meat products originating from North Wales. The low-level counting systems consist of high-efficiency, hyperpure germanium detectors with Compton suppression shields enclosed in lead castles to reduce background radiation. All systems are efficiency calibrated so that radionuclide concentrations can be determined.

A recent study of beef and lamb products in this laboratory revealed the presence of the radionuclide ^{110m}Ag ($T_{1/2}=250$ days) in both beef and lamb liver but not in other meats from the same animals. This radionuclide is readily identified by the presence of a γ -ray line of 658 keV alongside the characteristic 662 keV line of ^{137}Cs , although several other γ -rays including those at 447, 620, 707, 764, 885 and 937

keV provide an unambiguous identification of ^{110m}Ag .

A relevant portion of the γ -ray spectrum shown in Fig. 1 indicates the excellence of the detection system with its low background contribution and the clear resolution of the 658 and 662 keV lines.

Although the levels of ^{110m}Ag are small in both the beef and lamb liver, this radionuclide has not been observed in any other samples measured in this laboratory including air filters, rainwater and milk measured on 5–9 May 1986, nor in other meats and from ecological studies since.

In the beef liver we observed 23 ± 1 Bq kg^{-1} of ^{110m}Ag , 19 ± 1 Bq kg^{-1} of ^{134}Cs and 21 ± 1 Bq kg^{-1} of ^{137}Cs . In lamb liver we observed 74 ± 1 Bq kg^{-1} of ^{110m}Ag , 30 ± 1 Bq kg^{-1} of ^{134}Cs and 48 ± 1 Bq kg^{-1} of ^{137}Cs .

The radionuclide ^{110m}Ag is unlikely to arise as a fission product of ^{235}U . $A=110$ is near the minimum of the saddle point in the asymmetric mass distribution arising from the fission of ^{235}U by thermal neutrons. Also, a neutron-rich mass chain of $A=110$ produced in fission would terminate after successive β -decays at the stable nucleus ^{110}Pd before ^{110m}Ag could be reached.

A more plausible explanation is that the ^{110m}Ag is generated via the $^{110}\text{Cd}(n,p)^{110m}\text{Ag}$ reaction. Cadmium has been widely used in nuclear reactors as a neutron absorber for control purposes, however, following the Chernobyl reactor accident, cadmium may well have been used to try to control self-sustaining fission after the core melt-down.

As liver acts as a filter for large particles in the blood it is perhaps surprising that

^{110m}Ag has not previously been observed in dust from air filters although the fallout on Merseyside was slight compared to other parts of the country.

Further study of the lungs, liver and kidneys of animals reared in the areas of heavy fallout may provide yet more information concerning the Chernobyl accident.

G. D. JONES

P. D. FORSYTH

P. G. APPLEBY

Oliver Lodge Laboratory,
University of Liverpool,
Liverpool L69 3BX, UK

Characteristics of Chernobyl radioactivity in Tennessee

SIR—Fission product radioactivity from the Chernobyl reactor accident was first detected in air samples at Oak Ridge, Tennessee on 10 May 1986. The aerodynamic sizes of the aerosol-associated fission products were evaluated in four measurements made during the period 7 May to 13 June, using Sierra high-volume cascade impactors and low-background intrinsic germanium coaxial and well detectors. Additional air and precipitation samples were also obtained over this period. The Chernobyl activity was mostly sub-micrometre in size, and the particle size increased significantly over the measurement period. Only a small fraction of the aerosol ^{131}I was soluble in CHCl_3 . Both $^{134,137}\text{Cs}$ and ^{106}Ru were less soluble than natural radioactivity, indicating some association with aerosols produced by the accident.

Two distinct phases in airborne radioactivity were evident in our measurements. The first phase lasted from 10 to 17 May and was characterized by a $^{137}\text{Cs}/^{103}\text{Ru}$ activity ratio of ~ 1.5 , resembling the airborne fission products and deposition reported in Finland on the 28–29 April¹. The second phase began on 18 May, when precipitation from a convective storm yielded much higher ^{103}Ru and ^{140}Ba activities, relative to ^{137}Cs , than previously found in air or precipitation. This resulted in $^{137}\text{Cs}/^{103}\text{Ru}$ ratios of <0.8 , similar to that calculated from British data². Subsequent air and precipitation samples showed that this lower $^{137}\text{Cs}/^{103}\text{Ru}$ ratio persisted until 13 June. Table 1 summarizes these data.

The change in $^{137}\text{Cs}/^{103}\text{Ru}$ ratio on 18 May is interpreted as reflecting separate releases from the reactor which were meteorologically and/or temporally isolated. Two separate maxima in radioactivity were reported in Sweden, one on 28 April and one in the first week of May³. The highest measured air concentration of ^{103}Ru , 3.7 mBq m^{-3} , occurred in the 20–23 May sample. By 3 June, ^{103}Ru had declined to 0.1 mBq m^{-3} . We have also measured the amount of aerosol ^{131}I and its chemical

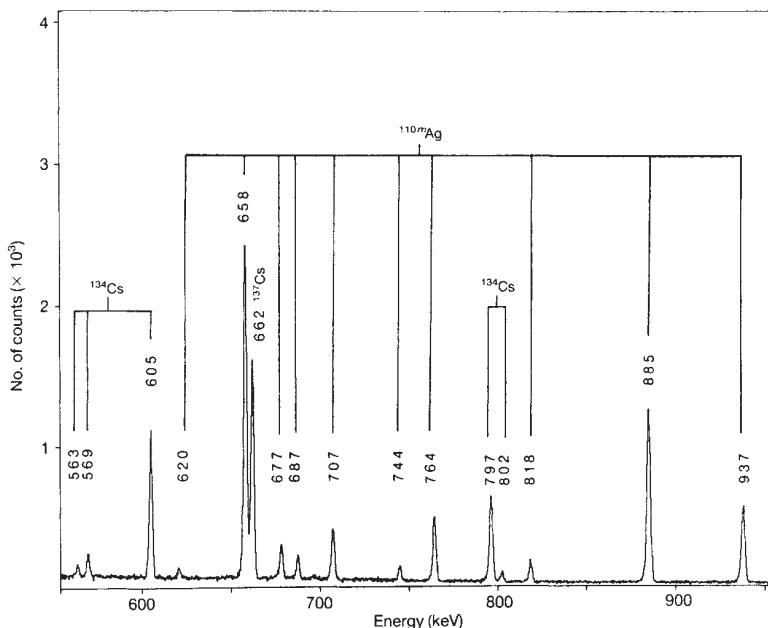


Fig. 1 A partial γ -ray spectrum arising from a sample of liver from a Welsh lamb. The count time was 57,700s. The lines are labelled by their energies in keV and by the originating radionuclides ^{110m}Ag , ^{134}Cs and ^{137}Cs .

Table 1 ^{137}Cs : fission product ratios (calculated as of 16 May 1986) measured in air (aerosol fraction) or precipitation

Date	Sample	Fission product				
		^{131}I	^{103}Ru	^{134}Cs	^{106}Ru	$^{140}\text{Ba}^*$
7–16 May	Impactor (air) ⁺	0.89	1.51	1.87	ND	28
13 May	Precipitation	0.14	1.53	ND	ND	ND
17 May	Air	1.58	1.50	1.94	ND	ND
18 May	Precipitation	0.09	0.76	2.43	ND	ND
20–23 May	Impactor (air) ⁺	1.00	0.56	2.20	4.60	2.54
23–27 May	Precipitation	0.13	0.72	1.45	2.69	2.04
28–29 May	Air	ND	0.50	1.78	ND	ND
2–3 June	Air	ND	0.50	1.79	ND	ND
6–13 June	Impactor (air) ⁺	ND	0.45	1.67	ND	ND
28–29 April	Helsinki air [‡]	1.0‡	2.5	1.7	—	1.78
29 April	Helsinki deposition [‡]	0.9	1.8	1.8	—	3.9
2 May	Chilton air [‡]	—	0.81	2.4	1.7	6.8

* ^{140}Ba or derived from ^{140}La , assuming secular equilibrium.

† Based on four stages from two cascade impactors.

‡ Aerosol ratio calculated from total ^{131}I , assuming 15% aerosol¹.

ND, not detected or measured with >25% uncertainty.

Table 2 Activity median aerodynamic diameters (μm) or Chernobyl fission products and cosmogenic ^7Be measured at Oak Ridge

Date	^{131}I	^{137}Cs	^{134}Cs	^{103}Ru	^{106}Ru	^{140}Ba	^7Be
7–16 May	0.37	0.40	0.40	0.37	ND	ND	0.38
20–23 May	0.32	0.48	0.43	0.44	0.44	0.45	0.39
30 May–3 June	ND	0.67	0.66	0.60	ND	ND	0.48
6–13 June	ND	0.72	0.71	0.62	ND	ND	0.36

ND, not detected or measured with >25% uncertainty.

composition. One of the cascade impactors operated from 7 to 16 May had a charcoal-impregnated final filter instead of the normal glass-fibre filter. Based on the ^{131}I distributions on both charcoal-equipped and normal impactors, we have estimated that $\leq 40\%$ of the airborne ^{131}I was aerosol-associated. The actual fraction may have been smaller, as the efficiency of the filter for trapping gaseous ^{131}I at high flow rate is not known. Both the Finnish¹ and Swedish³ reports indicated that $\sim 15\%$ of the ^{131}I measured on 28 May was gaseous. If that fraction is applied to the Finnish data¹, the resulting ratio (1.0) agrees well with our aerosol $^{137}\text{Cs}/^{131}\text{I}$ ratio from the 7–16 May measurement (0.89; see Table 1). Despite these qualifications, it is apparent that very little gas-to-aerosol transformation of ^{131}I occurred during its transit to the United States.

Aerosol ^{131}I speciation was examined using methods previously applied to weapons fallout⁴. The $<0.41\text{-}\mu\text{m}$ aerosol ^{131}I fraction, in both the 7–13 and 20–23 May measurements, consisted of $\sim 20\%$ I_2 or I (or other CHCl_3 -soluble species). No IO_3^- was detected, in contrast to what was found for weapons-derived ^{131}I (ref. 4). The $0.41\text{--}0.73\text{-}\mu\text{m}$ fractions from these measurements showed similar

CHCl_3 -soluble fractions.

Table 2 lists the mean aerodynamic sizes of aerosol-associated fission products reaching Tennessee. The fission product sizes increased considerably during the measurement period. By contrast, cosmogenic ^7Be , which is initially associated with very small aerosols, is continually being added to the aerosol population. In the first two impactor measurements, $\leq 2\%$ of the fission products were found to be larger than $2.1\text{ }\mu\text{m}$, strongly suggesting the absence of any significant mechanical suspension of reactor material into the free troposphere. The behaviour and sizes of the fission products are consistent with the long-range transport of isotopes released by volatilization, which condensed on or with ambient or accident-produced aerosols. We did not find ^{90}Zr or rare-earth fission products in our samples.

About 27% of the $^{134,137}\text{Cs}$ and 46% of the Ru was not rapidly dissolved in cold 2M HCl . This behaviour is very different from what is observed for natural ^7Be , ^{210}Pb or ^{212}Pb , which become associated with natural aerosols by surface condensation and are quite soluble. This poor solubility may indicate that some of the aerosol-associated fission products were

incorporated into reactor-derived materials which condensed in the hot plume leaving the reactor core. Autoradiographic evidence of such aerosols has been reported⁵.

This work was sponsored by the Office of Health and Environmental Research, US Department of Energy, under contract DE-ACO5-84OR21400 with Martin Marietta Energy Systems, Inc.

ERNEST A. BONDIETTI

J.N. BRANTLEY

Environmental Sciences Division,
Oak Ridge National Laboratory,
Oak Ridge, Tennessee 37831, USA

1. Savolainen, I. *Finnish Cent. Radiat. nucl. Safety Rep. No. STUK-B-VALO 44* (1986).

2. Fry, F.A., Clarke, R.H. & O'Riordan, M.C. *Nature* **321**, 193–195 (1986).

3. Devell, L. *et al. Nature* **321**, 192–193 (1986).

4. Perkins, R.W. *Health Phys.* **9**, 1113–1122 (1963).

Human lipocortin similar to ras gene products

SIR—Wallner *et al.*¹ recently reported the sequence of a human lipocortin complementary DNA clone. We divided the predicted amino acid sequence of lipocortin into several fragments and, using the similarity matrix comparison program FASTP², the fragments were searched against each other in order to locate any regions of internal homology. Analysis of the computer alignments of the fragments, combined with visual alignment using the nucleic acid sequence as a guide, revealed that, except for the first 43 amino acids, the protein sequence is composed of four repeats (Fig. 1). Statistical evaluations of all pairwise alignments of the repeats indicate a strong homology between most of them (Table 1). Com-

Table 1 Similarly scores (z) of the alignments of the lipocortin repeats

A	B	z(s.d.)
I + II	III + IV	24.4
I	II	11.8
I	III	8.5
I	IV	11.4
II	III	9.3
II	IV	19.4
III	IV	3.7

The FASTP program first identifies regions of sequence similarity, and then scores the aligned identical and differing residues in those regions by means of an amino acid replaceability matrix. The statistical significance of the alignment score between the repeat listed under A and the repeat under B is expressed as the similarity score (z), which is the number of standard deviations (s.d.) that the alignment score differs from the mean of the alignment scores obtained when A was compared to 100 random shuffles of sequence B. I + II and III + IV are amino acids 44–199 and 200–346, respectively. Scores between $z = 3$ and $z = 6$ are considered possibly significant, between 6 and 10 probably significant, and greater than 10 significant².

I 44 PSSDVAALHKA-IMVKGVD EATIIDILTKRNAQRQIKAAYLQETGKPLD LTKKALTHLEEVLLAKLTP 115
 II 116 AQPDADELRAA-MKGLGTDEDTLIEILASTNKEIKDINKVYREELKKRLKDLTISITSGDFNRALLSLAKGDRSEDFGVNEDL 198
 III 199 ADSARALDIEYAGERRKGTDVNVFNTILTKTSYPLRDFVYKITYKSHGMDKNVLDLTKGDIKCLTAIVCKATSK 275
 IV 276 AAF-AEKLHQA-MKGVGTTRHKLIRIMVSEGLIMMFIQKMYGKISQALIDELTGQYDKELKVALCGN 346 (COOH)

Fig. 1 Alignment of the four repeats in lipocortin. The complete predicted amino acid sequence is shown except for the amino-terminal 43 residues. Two asterisks indicate positions where three or four identical matches occur, and single asterisks where one or two pairs of residues are the same.

Lipocortin	78	QQIKAAALYETGKPLDETLKKAALTGHLVEEVLALLKTPAQFADELRAAMKGLGTDEDTLIEILASRTNKEIKRDINRVYR	157
C-K-ras2a	22	QLIQNHVEVDYPTIEISGRKQVVIDGETDCLDLDTAGQEYSAMRDQY--MRTGEGFLC-VFAINNTKSFEDIHH--YR	197
Lipocortin	158	EEKRLKDLADITSSTGDFRNALLSLAGKRSSEDFGVNEDLADSDARALYEAGERRRGTDVNVFNTIL--TTRTSYQLRRV	236
C-K-ras2a	98	EQKRLK--VKD--SEDEVPMVLGNKCDLP--SKTVTDQAQDLRLSYIPTSIAKTRQGVEDAFYLVIREIKQY--RLKKI	171

Fig. 2 Alignment of the predicted amino acid sequences of human lipocortin¹ and the human proto-oncogene c-K *ras2a*¹². Identities are indicated by asterisks and favoured amino-acid substitutions by dots. Favoured amino-acid substitutions¹³ are defined by pairs of residues belonging to one of the following groups: S/T, A/G and P/N, D/E and Q/R, K, and H/M, I/L and V/F, Y and W.

parison of the nucleotide sequences of the four repeats, aligned as in Fig. 1, revealed 41% homology on average, with repeat II being 50% homologous to repeat IV. This suggests that the four repeats evolved from a common sequence by means of gene duplication events. To what extent these structural units are congruent with the functionally active fragments³⁻⁷ of lipocortin remains to be determined.

After original submission of this correspondence, Kretsinger and Creutz⁸ described three regions in lipocortin which are homologous to a 17 amino acid consensus sequence⁹ from nine peptide fragments of the four related proteins P36, pII, calelectrin and endonexin. The homologous areas of lipocortin that were identified by these authors are amino acids 56 to 72 in repeat I and the corresponding areas of repeats II and IV (Fig. 1).

To find additional homologies to other proteins, we searched the predicted amino acid sequence of lipocortin against 3,061 sequences in the Protein Sequence Database of the Protein Identification Resource (PIR)¹⁰. The FASTP program revealed a strong homology between lipocortin and the mammalian *ras* proto-oncogene products, which are thought to interact with phospholipase C in a G-protein like fashion¹¹. The similarity score ($z = 8.1$ standard deviations) of the alignment between lipocortin and c-K-*ras*2a (Fig.2) is higher than the scores found by comparing *ras* to any of the known G-protein sequences.

TERRY Z. MUNN
GABRIELE I. MUES

*Department of Psychiatry,
University of Texas Health
Science Center at Dallas,
Dallas, Texas 75235, USA*

1. Wallner, B.P. *et al.* *Nature* **320**, 77 (1986).
2. Lipman, D.J. & Pearson, W.R. *Science* **27**, 1435 (1985).

3. Blackwell, G.J. *et al.* *Nature* **287**, 147 (1980).
4. Hirata, F. *J. biol. Chem.* **256**, 7730 (1981).
5. Cloix, J.F. *et al. Br. J. Pharmac.* **79**, 313 (1983).
6. DiRosa, M. *et al. Prostaglandins* **28**, 441 (1984).
7. Errasfa, N. *et al. Biochem. biophys. Res. Commun.* **847**, 247 (1985).
8. Kretsinger, R.H. & Creutz, C.E. *Nature* **320**, 573 (1986).
9. Geisow, M.J. *et al. Nature* **320**, 636 (1986).
10. Barker, W.C. *et al. Report of the National Biomedical Research Foundation* (Georgetown University, Washington DC, 1984).
11. Fleischman, L.F. *et al. Science* **231**, 407 (1986).
12. McGrath, J.P. *et al. Nature* **304**, 501 (1983).
13. Dayhoff, M.P. *et al. in Atlas of Protein Sequence and Structure* Vol.5, Suppl. 3 (ed. Dayhoff, M.O.) 345-352 (National Biomedical Research Foundation, Washington DC, 1978).

Evolutionary 'classics' may self-destruct

SIR—While respectively amused and bemused to see myself described as young, at 30, and as an entomologist, never having studied entomology, I feel compelled to comment on the further plethora of misinformation contained in Sibatani's recent letter¹, which attempted to defend the theory of evolution of Japanese scientist Imanishi against criticisms by Beverly Halstead². At the outset let me make it clear that, contrary to Sibatani's implication in his final paragraph, my views on Imanishi's theory of evolution have not changed as a result of working with Kazumi Tanida nor do I retract any of the advice I gave to Beverly Halstead during his stay in Japan.

As the author and I have never met, the complete misrepresentation of my standpoint and persona are, perhaps, understandable. Unfortunately, the conclusions and arguments forwarded seem similarly based on a complete unacquaintance with the subject. Confusion regarding the concepts in question is evident, as typified by the paralogistic statement "... preferences may well *partially* overlap and are thus unlikely to explain the *almost* total segregation...observed...", and is also ap-

parent in the superficial dismissal of Halstead's supporting references³⁻⁷. An informed reading of refs 3-5 will reveal them to demonstrate that groups of aquatic insects aggregate due to certain environmental and physical factors, and not due to any inherent species "proto-identity", as Imanishi would contend.

Reference 4 shows how selection of a physical variable by mayfly nymphs can "contribute significantly to the occurrence of highest densities on certain substrates and explain density shifts noted by other investigators in the field", thus adequately explaining Imanishi's wholly anecdotal field observations.

As only 0.3% of ref. 6 is concerned with "temporal segregation", it can hardly be said to "deal" with that topic, and in fact shows field and laboratory evidence of the importance of physical factors in the distribution of two sympatric trichopteran larvae. The extremely strictureed rebuff, that physical preferences are accepted but may well overlap, seems to have been formulated in splendid detachment from harsh ecological reality. Rivers are not clinal continua and habitat quality can vary both temporally and spatially. Although various individual factors can affect life history parameters and maximize individual fitness⁸, life history and distribution patterns usually represent a compromise between an interacting complex of biotic and environmental factors.

Competition itself is an interaction⁹, and so the attempted tautological evasion of the significance of the contents of ref. 7 fails, but does emphasize the need for a standardized definition of competition and its occurrence¹⁰.

As for the quoted confirmatory evidence in the net-spinning Hydropsychidae¹¹, I feel that samples taken on only three occasions, two of which were 7 years apart, do not form an adequate logistic base for such conclusions, and the apparent loss of one species from the site during the sample period may actually provide evidence for competitive displacement or exclusion. Furthermore, in this family, the documented sequential downstream species-replacement, sometimes overlapping, sometimes disjunct, is well known¹², detailed studies having revealed it to be mainly related to variations in the physical factor gradients down the length of a watercourse¹²⁻¹⁵, while at the microhabitat level, species coexisting differed in their biotic and physical factor preferences, results agreeing with theoretical arguments based on resource partitioning¹⁶ and depressibility^{17,18}. The regular spacing in some sites indicates territoriality, itself a form of intraspecific competition¹⁹. Larvae may fight for net sites and also possess an ultrasonic stridulatory mechanism which is used agonistically to repel intruders and to defend the site²⁰. In this family therefore, the evidence for

inter- and intra-specific competition is overwhelming and the distribution patterns are easily explained in terms of orthodox neo-darwinian theory. However, although a growing body of research has shown competition to play a critical role in determining habitat selection, survivorship and, ultimately, Darwinian fitness in some stream insects²¹, in other species, for example, mayfly larvae, little interspecific interaction was evident under certain conditions^{22,23}. There exist an immense variety of possible interspecific interactions, and dangers are inherent in overemphasizing any single aspect, and in overgeneralizing from observations taken either over a limited period, in a limited situation, or on a restricted range of species.

Scientific evidence should be based upon sound scientific data, and in this respect Imanishi's theory is found to be seriously wanting. Viewed from the ethereal realm of the philosopher, biological phenomena often differ in quite fundamental aspects than when viewed from the quantitative world of the scientist. Thus, Imanishi may have suffered the singular misfortune of basing his evolutionary theory upon non-quantitative observations made in a situation where little or no interaction occurred during the observation period. Such vague and essentially untestable theories are always difficult to disprove, but quantitative evidence usually eventually supercedes bad philosophy. As Wordsworth so eloquently phrased it, "Our meddling intellect misshapes the beauteous form of things", yet accuracy is the goal we seek. In this respect, the proposed translations of Imanishi's 'classics' may prove productive since, in addition to being "unrefined and crude"²⁴, subjective analysis, coupled with the quantitative evidence from a growing body of research may reveal them to contain the ingredients of their own implausibility.

ANDREW ROSSITER

Laboratory of Animal Ecology,
Kyoto University,
Sakyo-ku,
Kyoto 606, Japan

1. Sibatani, A. *Nature* **320**, 492 (1986).
2. Halstead, L.B. *Nature* **317**, 587 (1985).
3. Barber, E.W. & Kevern, N.R. *Hydrobiologia* **43**, 53–75 (1973).
4. Wright, L.L. & Mattice, J.S. *Aquatic Insects* **3**, 13–24 (1981).
5. Hawkins, C.P. & Sedell, J.R. *Ecology* **62**, 387–397 (1981).
6. Cummins, K.W. *Ecol. Monogr.* **34**, 271–295 (1964).
7. Schoener, T.W. *Am. Nat.* **122**, 240–285 (1983).
8. Sweeney, B.W. in *Ecology of Aquatic Insects* (eds Resh, V.H. & Rosenberg, D.M.) (Praeger, New York, 1984).
9. Krebs, C.J. *Ecology: The Experimental Analysis of Distribution and Abundance* (Harper & Row, New York, 1978).
10. Ferson, S. *et al. Am. Nat.* **127**, 571–576 (1986).
11. Tanida, K. *Physiol. Ecol. Jap.* **21**, 115–130 (1984).
12. Edington, J.M. & Hildrew, A.G. *Caseless Caddis Larvae of the British Isles* (FBA Pub. No. 43, 1981).
13. Edington, J.M. & Hildrew, A.G. *Verh. int. Verein. theor. angew. Limnol.* **18**, 1549–1558 (1979).
14. Boon, P.J. *Hydrobiologia* **57**, 167–174 (1978).
15. Hildrew, A.G. & Edington, J.M. *J. Anim. Ecol.* **48**, 557–576 (1979).
16. May, R.M. & MacArthur, R.H. *Proc. natn. Acad. Sci. U.S.A.* **69**, 1109–1113 (1972).

17. Royama, T. *J. Anim. Ecol.* **42**, 693–726 (1974).
18. Hassell, M.P. & May, R.M. *J. Anim. Ecol.* **42**, 693–726 (1974).
19. Odum, E.P. *Fundamentals of Ecology* (Saunders, Philadelphia, 1971).
20. Jansson, A. & Vuoristo, T. *Behaviour* **71**, 168–186 (1979).
21. Hart, D.D. in *Stream Ecology: Application and Testing of General Ecological Theory* (eds Barnes, J.R. & Minshall, G.W.) (Plenum, New York, 1983).
22. Peckarsky, B.L. & Dodson, S.I. *Ecology* **61**, 1275–1282 (1980).
23. Corkum, L.D. & Clifford, H.F. in *Advances in Ephemeropteran Biology* (eds Flannagan, J.F. & Marshall, K.E.) (Plenum, New York, 1980).

Recent discoveries of a supermass in the Universe

SIR—A pair of quasars with very similar redshifts was recently interpreted as a pair of images produced by a gravitational lens of enormous mass¹. However, I and the other astronomers who originally discovered the quasars^{2,3} pointed out that this pair belonged to a tight association of five quasars on the sky. The spatial grouping alone was equivalent to a 20 sigma density enhancement. The similarities in redshift of these quasars — four were at redshifts $z=0.9$, 1.0, 1.0 and 1.1 — had in addition only a chance of $\approx 10^{-4}$ of being accidental. Later it was shown⁴ that all the densest concentrations of quasars known in the sky had this concentration of redshifts around $z \approx 1$.

For gravitational lens advocates to now fasten on certain aspects of only two objects in these physical groups which fit their interpretation — and totally ignore the other information which does not fit — is tantamount to distorting the scientific data.

The enormity of the derived mass also makes it clear that there is no result which is sufficiently absurd to force rejection of the original assumption. A scientific theory must be capable of being disproved which apparently this is not.

A fascinating sequel quickly emerges when it is shown that the two supposed images of the same object are not identical in the red⁵. The simple conclusion which follows is that these two quasars, which have almost identical spectra, are really two separate objects. But then what happens to the arguments for the original gravitational lens? (called Q0957+561; see refs 6–8).

The argument that it had to be a gravitational lens rested on the claim that there was no other explanation for how the spectra could be so similar. (Actually one was considerably redder than the other.) What is the evidence now for gravitational lenses?

In general we should not disguise the fact that gravitational lens theories are really just another variation of 'hidden mass' or 'dark matter' hypotheses. Just as undetected (or undetectable) matter is supposed to close the Universe, so unmeasured mass in clusters of galaxies is supposed to explain large redshift dispersions

and unseen and implausibly distributed mass is supposed to explain flat rotation curves in spiral galaxies. Until scientific detection and measures of these hypothesized entities is made they should be recognized as really just statements that some part of our current physical laws are contradicted by the observations.

HALTON ARP

Max-Planck-Institut für Astrophysik,
Karl-Schwarzschild-Str. 1,
D-8046 Garching b. München, FRG

1. Turner, E.L. *et al. Nature* **321**, 142 (1986).
2. Hazard, C., Arp, H. & Morton, D.C. *Nature* **282**, 271 (1979).
3. Arp, H. & Hazard, C. *Astrophys. J.* **240**, 726 (1980).
4. Arp, H., 307–352 (Universite de Liege, Institut d'Astrophysique, 1983); *Quasars Redshifts and Controversies*. (Interstellar Media, Berkeley, California, 1986).
5. Shaver, P. & Christiansi, S., *Nature* in press.
6. Walsh, D., Carswell, R.F. & Weymann, R.J. *Nature* **279**, 381 (1979).
7. Weymann, R.J., *et al. Astrophys. J.* **233**, L43 (1979).
8. Young, P., Gunn, J.E., Kristian, J., Oke, J.B. & Westphal, J.A. *Astrophys. J.* **241**, 507–520 (1980).

Espresso coffee emporia take note

SIR—Apfel and Davy's report (*Nature* **321**, 658; 1986) of the superheating of water by microwave ovens is totally consistent with empirical evidence obtained while using our laboratory microwave oven for preparing hot drinks (a practice sure to get us into hot water with our safety officer). I had placed a cup of tap water in the microwave and set it to boil but at first was prevented from making the drink by an experiment that required attention. On the third occasion of activating the oven, I responded to the timer bell by rushing to the oven, removing the cup and stirring my hot chocolate powder into the water which promptly erupted into a foaming cascade.

An interpretation of this dramatic event was that the water had become superheated and that the chocolate powder simply provided nucleation sites for the water to boil. In the light of Apfel and Davy's quantitative study of the superheating of water in microwave ovens it would seem that the explanation was correct. As well as offering supportive evidence and perhaps a cautionary tale, we think that a practical application of this phenomenon with coffee might be the production of energy-efficient espressos!

ANTHONY PARSONS

Department of Biology,
University of York,
York YO1 5DD,
UK

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a rather technical character which are not provoked by articles or letters previously published.

Tensions in science

David Dickson

Scientists and Journalists: Reporting Science as News.

Edited by Sharon M. Friedman, Sharon Dunwoody and Carol L. Rogers.

Free Press/Collier Macmillan: 1986. Pp.333. \$24.95, £25.

SHORTLY after taking over as the new administrator of the US National Aeronautics and Space Administration at the end of May, Dr James C. Fletcher delivered a stinging attack on the way that the US news media had covered the Challenger accident and its aftermath. Speaking to an audience of aerospace executives, he claimed that the media had acquired a "deep and unwarranted suspicion of NASA", reflected in its highly critical investigations of the agency's management, and that the resulting coverage could do "irreparable damage" both to the space programme and to the nation.

As several journalists were quick to point out, there was a certain irony in Fletcher's remarks. For NASA has always depended, perhaps more than any other US government agency, on heavy media coverage of its activities to generate the public enthusiasm needed to justify a continued investment of funds. Furthermore, studies of the agency's relations with the media in the 1960s suggest that a "myth of invincibility" had been deliberately developed, sustained by a lack of critical reporting, and became a factor in creating the conditions leading to the Apollo fire in 1969 in which three astronauts died.

Fletcher's remarks, however, highlight the ambiguity that surrounds public perceptions of science and technology in modern culture. Both are viewed as sources of trust and authority, but at the same time both are seen as containing threats that undermine the assumptions on which this trust is based. The ambiguity is reflected in two very different aspects of science journalism, and the tension between them emerges from a close reading of *Scientists and Journalists*, a collection of essays by working journalists, media sociologists and practising scientists.

The core of the book is a series of presentations delivered to a symposium organized by the American Association for the Advancement of Science. To these have been added further contributions which, while not exhaustive in their coverage (other views are referred to in a valuable annotated bibliography), nevertheless highlight many of the complexities of what is often considered a one-dimensional debate. Furthermore, although *Scientists and Journalists* is written primarily for an American readership, the perspectives on display are equally applicable to the practice of science journalism

in other countries, Britain in particular.

Several articles in the volume provide detailed accounts of the way that the different actors concerned with the communication of information about science see their professional responsibilities. An articulate statement of the journalist's role is provided by Christine Russell, a science writer on the *Washington Post*, while that of the public relations officer is described by Carol Rogers, who is responsible for such activities at the AAAS.

Other articles give insights either into the social factors that influence how science journalists operate as a community, or into the challenges to prevailing approaches to writing about science. In the first category Sharon Dunwoody, who teaches journalism at the University of Wisconsin, describes the existence of a "science writing inner club", made up primarily of experienced journalists from the main news media in the United States, and the way this "club" can influence what is treated as news by other science journalists. In the second category, Jon Franklin, a Pulitzer prize winner who writes for the *Baltimore Evening Sun*, tells how he has developed a highly-personal, narrative style of science writing. A closely annotated version of one of his essays, describing a surgeon's unsuccessful attempt to remove a brain tumour, provides useful hints for conveying scientific information in a readable and absorbing manner.

Taken together, these somewhat heterogeneous contributions give a valuable overview of many of the issues that are

likely to confront either journalists or scientists as each side struggles to make sense of the other.

At the same time, several of the authors address, sometimes only in passing, the question "where do we go from here?". And it is on this point that the tensions and disagreements arising from the ambiguities described at the beginning of this review begin to appear. For example, implicit in many of the contributions, as indeed in much of the current debate about the "public understanding of science" on both sides of the Atlantic, is that better science journalism will result from closer collaboration between scientists and journalists. Thus Dunwoody suggests the need to develop a "shared culture" of reporters and scientists which, she considers, has already developed from virtual non-existence to "a kind of fragile, neophyte state". Rae Goodell, however, a teacher of science writing at the Massachusetts Institute of Technology, uses an analysis of the recent controversy over the safety of recombinant DNA experiments to argue that an excessively close relationship can enable scientists to manipulate the media to their own advantage.

Elsewhere, divergent views about the social function of science journalism reveal themselves as reflections of broader differences over desirable forms of social decision-making. Thus Jon D. Miller suggests that the main target for news about science, particularly those aspects relevant to policy issues, should be those at or near the top of the political tree. Under normal circumstances, he says, "I don't think any useful purpose would be served by directing policy-relevant information to the nonattentive segment of the public" (those he defines as lacking either a high degree of interest in science policy, or a "functional understanding of the process of terminology of science"). In contrast, David W. Crisp, a reporter for a local newspaper in Texas

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

NASA has depended on the media to generate public enthusiasm for its activities.

who won an award for his coverage of the federal government's proposal to build a nuclear waste dump near his town, describes how under certain conditions the science journalist can be thrust into the role of providing crucial information to those who might otherwise be deprived of it — or who distrust other sources.

Sadly, but perhaps wisely, the editors have not attempted to pull together the wide variety of points of view represented in the book. If they had done, they would have found considerable evidence to support the argument that a closer relationship between working scientists and journalists would undoubtedly be more comfortable for science. But whether it would necessarily be equally beneficial to the non-scientific community is less clear. As William Bennett, editor of the *Harvard Medical School Health Letter*, points out, the journalist's role of "informed sceptic" is vital to the health of a science-

based democracy. And this role is never a straightforward one; in science, even more than in politics or the arts, the definition of "informed" is often open to dispute.

Nevertheless scientists who are quick to attack critical reporting of their work, or that of their colleagues, should consider a thoughtful comment from David Crisp: "Good scientists and good journalists," says Crisp, "have a great deal in common: an ingrained skepticism toward established dogma, conventional wisdom and the tyranny of commonsense; an eye for accuracy and detail; the ability to deal impartially with facts that don't fit the theory; and contempt for the pernicious theory that all passionately held ideas are of equal value". □

David Dickson, *Le Billehou, 91190 Gif-sur-Yvette, France, is the European correspondent for Science, and the author of The New Politics of Science (Pantheon, 1984).*

Dynamic insights

J.R.L. Allen

Coastal and Estuarine Sediment Dynamics. By Keith R. Dyer. Wiley:1986. Pp.342. £36, \$63.35.

SEDIMENT transport dynamics is of great importance in several disciplines, each of which, in its traditional garb, approaches the subject from a special standpoint and with particular perceived needs. This is why it is so hard to write about it well. Sedimentologists and geomorphologists are best able to discuss the outcomes of sediment transport, namely what, in the form of sedimentary textures, structures and larger scale formations, has to be explained. From fluid dynamicists and rheologists come the necessary theoretical and experimental insights into the roles of fluid motion and particle mechanics in sediment transport. Engineers, with their emphasis on an empirical, pragmatic approach, are best equipped to provide from contemporary environments and from the laboratory many of the additional links necessary for the appreciation of sediment transport and its real effects in quantitative terms.

Writing about sediment dynamics as expressed in shallow-marine environments, Dr Dyer rightly attempts to blend these three traditional approaches — sadly, still a rare venture. The result is a lucidly written book which, if with limitations here and there on the descriptive and theoretical sides, and a singular lack of photographic illustrations, is up to date and largely a success.

The book begins with a brief survey of the subject and a sketch of sediment production and the geological context of

modern shallow-marine sedimentation. Next come discussions of sediment particles as such and the fluid flow; more emphasis could usefully have been placed on the diversity and significance for transport of particle shapes, and on mass transport currents in waves. Chapter 3 covers sediment entrainment, one-way transport in suspension and as bedload, and bedforms (inadequately characterized). The following brief chapter examines sediment movement under waves, but the resulting bedforms are again ill-characterized and the origin of wave ripples is not explored. The further treatment of sediment suspension includes useful accounts of stratification by suspended particles and the suspension profile under waves.

In Chapter 7 Dyer compares sediment transport-rate formulae and sketches the important but under-researched topic of transport beneath waves and waves combined with currents. Not surprisingly, given the author's background, the treatment of cohesive sediments is excellent, except for the surprising omission of the by no means negligible mass-movement processes. Estuarine and coastal sedimentation are well summarized, as are beach processes, but the important role of storms in causing unusual sediment transport in these environments is ignored. An extensive bibliography (c. 1,000 items) and a short index conclude the book. Readers should beware equations 3.6 and 11.17, incorrectly set, and a mis-spelling of Bernoulli. Regrettably, SI units are eschewed.

This book has its faults. But it broadens our appreciation of sediment dynamics in shallow-marine environments, and deserves to be read and used. □

J.R.L. Allen is a Professor in the Department of Geology, University of Reading, Whiteknights, Reading RG6 2AB, UK.

Framework for twistors

Abhay Ashtekar

Spinors and Space-Time. Vol. 2 Spinor and Twistor Methods in Space-Time Geometry. By R. Penrose and W. Rindler. Cambridge University Press:1986. Pp.501. £45, \$89.50.

TWISTORS were invented by Roger Penrose about 20 years ago. Since then he and his research group have worked on the subject steadily and have built it into a thriving branch of mathematical physics. The goal of the twistor programme is very ambitious — first to recast and then further develop all of known physics within a framework in which space-time itself is a derived, secondary object. Considering that relatively few individuals have been involved in the programme, the results to date are most impressive from a mathematical standpoint. Indeed, even if it should turn out that the ideas cannot be taken any further because of some intrinsic limitation, the fact that so much of the basic mathematical structure underlying relativistic physics can be constructed without a direct reference to space and time is, in itself, sufficient to make this endeavour fascinating. It is therefore regrettable that most mathematical physicists know virtually nothing about twistor theory. In part, this is because very little of the recent literature on the subject is accessible to non-specialists. The second volume of *Spinors and Space-Time* is intended to remedy this problem, and it does so very satisfactorily.

The authors first introduce the basic ideas of twistor theory, including some of the recent developments, and then go on to discuss a number of constructions of use in general relativity which are simplified and, in some cases, inspired by this theory. Since the relevant results and equations from Vol. 1 are recalled at the beginning of the book, readers familiar with 2-component spinors can go directly to Vol. 2.

The introduction to twistor theory is fairly detailed. Readers are first exposed to the non-local relation between the geometry of twistor space and of Minkowski space-time; they are then shown that the momentum angular-momentum structure of particles and fields can be coded in a natural way in the twistor picture; and, finally, they are presented with the description of the zero rest mass fields in terms of contour integrals and cohomology on the twistor space. It is remarkable that, although the relation between the twistor space and space-time is non-local, solutions to physically interesting local differential equations in space-

time can be represented in a natural way in twistor space. This interplay between cohomology on complex manifolds and solutions to differential equations on real manifolds is extremely interesting, because it brings together two otherwise unrelated branches of mathematics in an unexpected way. The outstanding application of these ideas to general relativity is the definition of quasi-local quantities which attempts to capture the notion of total energy-momentum and angular momentum associated with the region of space-time enclosed by a topological 2-sphere. The current status of this definition and its consequences is described in detail. In addition, several mathematical results on null congruences, on algebraic classification of curvature tensors and on asymptotic structure of space-time at null infinity, which are simplified and illuminated by the use of twistor methods and the spin coefficient techniques, are presented in varying degrees of detail and completeness.

This is primarily a work in mathematics rather than physics. The emphasis is on illustration of the power and elegance of certain techniques rather than on provision of a comprehensive treatment of topics that are of direct application in general relativity. Consequently, the discussion is not uniform; points dealing with twistor ideas and spin coefficient methods are treated thoroughly, while those requiring ideas from elsewhere are often mentioned only briefly.

Overall, there is a considerable difference between this volume and its companion. Volume 1 (reviewed in *Nature* 313, 607; 1985) appeared as a finished, definitive work that could be used in courses as well as by individual researchers. Volume 2, on the other hand, has a mixed flavour. Parts of it, such as the treatment of quasi-local mass, report the current state of the art. These parts have the style of lecture notes rather than that of a monograph. Furthermore, none of the chapters has an introductory section to map out the material that follows. Consequently, it is sometimes difficult to see where one is headed, and only those readers who are already familiar with more conventional treatment of topics such as gravitational radiation theory, and energy-momentum and angular momentum in general relativity, will be able to appreciate the simplifications and insights provided by twistor methods. Volume 2, therefore, is primarily for researchers who want to understand the spirit of twistor theory and for teachers of courses on general relativity who want to broaden their knowledge of the field. □

Abhay Ashtekar, a Professor of Physics at Syracuse University, New York, is currently at the Institute of Theoretical Physics, University of California, Santa Barbara, California 93106, USA.

Strategy for future weather

Nicholas A. Rupke

Climatic Change and World Affairs, Revised Edition. By Crispin Tickell. *Center for International Affairs, Harvard University/University Press of America: 1986. Pp.76. \$17.50; pbk \$7.25. Distributed in Britain by Eurospan, 3 Henrietta Street, London WC2E 8LU. Hbk £16.55; pbk £7.30.*

ERRATIC blocks and boulders, strewn across much of northern Europe and North America, reminded our ancestors of the waters of the biblical deluge. By around 1840, the Swiss naturalist Louis Agassiz concluded that these erratics had been put in place by gigantic ice sheets, and were evidence of a former Ice Age.



Heavy cloud in the southern hemisphere, as seen from Apollo 17. The photo extends from the Mediterranean to the Antarctic.

Thus began the scientific study of glacial climatic change in the geological past — change which, during the most recent, Quaternary Period of Earth history, has included half a dozen ice ages alternating with warmer interglacial stages. Simultaneously, however, our awareness of the vast extent of geological time has grown and, as a result, the historic reality of climatic upheaval has receded into the dim distance of “millions of years” before (or after) us.

Crispin Tickell’s “brilliant short book” (to quote from Lord Zuckerman’s introduction to the British edition) warns us, however, against a false sense of security and an *après nous le déluge* complacency. Our seemingly stable climate is in fact a vulnerable and changeable product of a

great many astronomical and terrestrial factors. Moreover, man himself has become a potentially destabilizing factor as a consequence of large-scale environmental mismanagement, such as the pollution of the air with carbon dioxide, aerosols and other industrial wastes which change the heat balance of our atmosphere.

Much has already been written about the threat to world climate — some authors demonstrating the scientific complexity and uncertainties of the issue, others giving free rein to doomsday predictions of imminent catastrophe in the form of a new and prolonged Arctic winter or, conversely, of the precipitous melting or surging of polar icecaps and a consequent rise in sea level. In this revised edition of a book first published in 1977, Tickell again succeeds in walking the tightrope between these two approaches: he presents an intelligible account of world climate in all its scientific complexity, but does so without letting the reader

lose a sense of the reality and potential immediacy of a calamitous deterioration. After all, only a relatively small change of a few degrees Celsius in mean global temperature would suffice to trigger an irreversible feedback process of glacial expansion or retreat at the poles.

The author (who is Permanent Secretary of the Overseas Development Administration in London) has examined the subject with the eyes of both science and international affairs. In a “call for action” he forcefully argues for the need to get some form of international agreement on how to cope with future climatic crises. More specifically, Tickell proposes that administrative responsibility be given to the World Climate

Program (adopted in 1979 by the World Meteorological Organization) as the custodian of a series of climatic agreements. These should cover all major experiments and actions (industrial and military) which affect the climate. Furthermore, an international code of good climatic behaviour should be adopted and enforced by means of various economic measures.

Beginners and specialists alike will find *Climatic Change and World Affairs* lucid and stimulating, particularly if they are interested in science policy and international cooperation in the study of weather and climate. □

Nicholas A. Rupke is a geologist and an historian of science at Wolfson College, Oxford OX2 6UD, UK.

Vegetative roots differ

Peter D. Moore

Studies on Plant Demography: A Festschrift for John L. Harper. Edited by James White. *Academic*. 1985. Pp. 416. Hbk £45, \$59.50; pbk £22, \$29.50.

The Population Structure of Vegetation. Handbook of Vegetation Science, Vol. 3. Edited by James White. *Dr W. Junk*. 1985. Pp. 669. Dfl. 310, \$97.50, £85.95.

THROUGHOUT the history of plant ecology there have been profound disagreements over the most realistic, or at least the most intellectually profitable, way of studying vegetation. Is vegetation simply a "fortuitous juxtaposition of plant individuals", as maintained by Gleason? Or is it an orderly assemblage of recurrent plant associations, as conceived by certain schools of phytosociology? Plant population biology is essentially the product of an individualistic way of thinking: it is concerned with the interactions between adjacent members of a population, their growth, reproduction and mortality. But the overview of vegetation which the phytosociologist seeks to present must still recognize populations of plants, perhaps co-evolved, as the elemental structures from which higher levels of organization are derived.

Two approaches to the subject are thus possible: we can begin with vegetation and work down towards the individual, or we can begin with the individual in its population and gradually extend towards the more complex interactions of the community. In the books under review we find the two schools of thought effectively and separately illustrated, which is a credit to James White who edited both volumes.

John Harper, to whom the first book is dedicated in celebration of his sixtieth birthday, is rightly regarded as the founding father of plant population biology, and his perspective on the subject is essentially Darwinian and analytical; a view of architecture that begins with the

bricks. This collection of papers by his past students amply demonstrates the value of such an approach. Take John Ogden's work on forests, for example. In this type of habitat the long life-history of the individual gives an impression of stability, but population studies reveal a constant turnover, a mosaic of recovery stages from past disturbances — what Henry Horn has described as a "pre-emptive crazy quilt" of vegetation.

Mobility, of course, is a special problem for plants, but one that must be overcome if the patchwork of opportunities in the environment is to be fully exploited. Noble illustrates the effectiveness of birds in transporting seeds even when the birds in question cannot fly, as in the case of the gut carriage of *Nitraria* by emus in Australia.

Turkington, on the other hand, expands on the value of vegetative growth by stolons in clover. Clones can spread extensively in this way (some ecologists even talk of such plants "foraging"), but though the propagation of vegetative clones provides a form of mobility it can lead to a general lack of genetic diversity. In Turkington's view, the input of new genotypes by somatic mutation and occasional germination of new individuals is adequate to provide this required genetic variability.

Studies such as these on clover can lead to confusion about what really constitutes an individual plant. One approach is to consider a plant as a collection of relatively independent units. This idea of modular growth, where each module is the product of a single meristem, is particularly attractive and is exemplified here by Michelle Jones's report of her work on silver birch. In the case of trees, the branching systems differ in their "modular dynamics" because they are influenced by neighbours and respond by the formation of varying crown geometries. The implications of such studies for forestry are further developed by Miguel Franco.

The total collection of papers provides a wide range of examples from current studies in plant population biology. Most of these have originated from Harper's work and ideas, and are now taking root and ramifying.

In *The Population Structure of Vegetation*, James White is at pains to encourage the recognition of the complementarity of the demographic and sociological approaches. But it is vegetation that is quite recognizably the starting point of most of the papers included here, rather than population studies on individual species. Perhaps the key to the difference in thinking lies in the paper by T.A. Rabotnov on the dynamics of plant coenotic populations. But what, we may be forgiven for asking, is a coenotic population? It is defined as the sum of the individuals of a species within a plant community (the phytocoenosis), and is a

concept arrived at by Rabotnov himself (the founding father of the Soviet plant coenotic population biology to whom this book is dedicated) as a result of the mental dissection of vegetation into its interacting components.

One of the consequences of this way of looking at population biology is the need to examine the demography of many different species at the same time; the detailed analysis of one species is of little value in isolation. Inevitably this leads to less-rigorous, less-numerical conclusions than typify the Harper school, but it also encourages the tendency to classify plants according to their interactive "strategies" as a means of simplifying community description. Some interesting proposals have resulted. The threefold classification of plants into "violents", "patients" and "explerents" by Ramenskii, for example, foreshadows the work of Grime, who also summarizes his current views in this volume. But such simple classification systems are not universally adopted by the contributors. Indeed, Grubb provides a detailed argument against the use of such models, demanding a much more critical attitude to the use of such terms as "stress", "tolerance", "dominance", "disturbance" and "competition", and calling for more attention to be given to the factors which permit coexistence among plants.

One prominent aspect of this book are the contributions of Soviet scientists, and Markov's paper on the use of permanent quadrats in the Soviet Union opens up for Westerners a useful window on Eastern ecology. Most of the Soviet papers are concerned with coenotic population research in steppic environments, but more occidental studies are also included, from chalk grasslands in the Netherlands to the cedar groves of the south-eastern United States. There are even some contributions that wander slightly from the "coenotic" theme, such as the demographic study of *Plantago major* and *P. lanceolata* by van der Aart. But even here the main emphasis is upon the variation in demography according to the plant's habitat — paths, hay fields, dunes, roadsides, pastures and so on — so the community remains the centre of interest.

Perhaps the most remarkable feature to emerge from these two books is the great difference in the starting point and the approach found in each and yet their tendency to generate lines of convergence. These lines will undoubtedly culminate in fruitful fusions, possibly under the diplomatic eye of James White, the sole common factor. Meanwhile, the balanced botanist must obviously have access to both. □

Peter D. Moore is Reader in Ecology in the Department of Biology, King's College London (KQC), 68 Half Moon Lane, London SE24 9JF, UK.

● R.K. Peet has edited *Plant Community Ecology: Papers in Honour of Robert H. Whittaker*, which contains papers written by Whittaker's past colleagues and students after his death in 1980. Rather than review Whittaker's own contributions, Peet chose to illustrate current applications of approaches Whittaker developed and to show recent advances which have grown from his pioneering work. The papers are arranged in four sections, representing areas of plant community ecology which were strongly influenced by Whittaker: Methods of community analysis, Analysis of gradients, Community dynamics and Species Diversity. Published by Dr W. Junk as Vol. 7 of the "Advances in Vegetation Science". Pp. 332. Dfl. 200, \$68, £55.50.

What to believe about miracles

from R.J. Berry

Differences of interpretation within the Church of England have stimulated a re-examination of the "miracles" as recounted in the Bible.

Two years ago I was one of 14 signatories of a letter to the *Times* about miracles¹. All of us were professors of science in British universities; six were Fellows of the Royal Society. We asserted: "It is not logically valid to use science as an argument against miracles. To believe that miracles cannot happen is as much an act of faith as to believe that they can happen. We gladly accept the virgin birth, the Gospel miracles, and the resurrection of Christ as historical events. . . . Miracles are unprecedented events. Whatever the current fashions in philosophy or the revelations of opinion polls may suggest, it is important to affirm that science (based as it is upon the observation of precedents) can have nothing to say on the subject. Its "laws" are only generalizations of our experience. . . ."

A leading article in *Nature*² accepting our statement on the nature of scientific laws, dissented from our conclusion about miracles, terming them "inexplicable and irreproducible phenomena (which) do not occur — a definition by exclusion of the concept. . . . the publication of Berry *et al.* provides a licence not merely for religious belief (which, on other grounds is unexceptionable) but for mischievous reports of all things paranormal, from ghosts to flying saucers".

Subsequent correspondents disagreed. For example, Clarke³ objected that "your concern not to license 'mischievous reports of all things paranormal' is no doubt motivated in the interest of scientific truth, but your strategy of defining away what you find unpalatable is the antithesis of scientific"; MacKay⁴ emphasized that "for the Christian believer, baseless credulity is a sin — a disservice to the God of truth. His belief in the resurrection does not stem from softness in his standards of evidence, but rather from the coherence with which (as he sees it) that particular unprecedented event fits into and makes sense of a great mass of data. . . . There is clearly no inconsistency in believing (with astonishment) in a unique event so well attested, while remaining unconvinced by spectacular stories of 'paranormal' occurrences that lack any comparable support".

The credibility of belief in miracles has resurfaced in a report of the Church of England bishops⁵. The *Times* commented: "Did the two key miracles at the centre of the Christian faith, the Virgin Birth and the Resurrection, really happen? . . . The exercise has established one thing clearly:

that belief in miracles, at least where they are central to the faith, is thoroughly intellectually respectable. . . ."⁶

It would be easy to decry the criteria or standards of proof accepted by the bishops, but their integrity is presumably not in doubt. It is more profitable to enquire whether miracles are really credible, and, if so, what are the circumstances where they might be expected.

Natural law

"In an earlier age, miracles would have been one of the strongest weapons in the armoury of apologetic. A man who did such things must at the very least have the power of God with him. Jesus himself is represented as using this argument when he said, 'If it is by the finger of God that I cast out demons then the kingdom of God has come upon you' (*Luke* 11:20). For us today, by one of those twists that make up intellectual history, miracles are rather an embarrassment. We are so impressed by the regularity of the world that any story which is full of strange happenings acquires an air of fairytale and invention"⁷. The historical twist referred to by Polkinghorne was an inevitable consequence of the separation of observation (or test) from interpretation, which is the essential feature of what we call science. Before the sixteenth century "how" and "why" questions were answered in much the same way: acorns fell to the ground so that new oaks might grow; rain came so crops might flourish and people feed; and so on. The realization that the same event could be interpreted in more than one way led to an emphasis on mechanism, and therefore on the uniformity and predictability of natural events, with a consequent restricting of divine activity to the ever-decreasing gaps in knowledge. God became unnecessary, except as a rationalization for the unexplainable⁸.

By the seventeenth century scientists were using the "laws of nature" in the modern sense, and the physical and (increasingly) the biological worlds were regarded as self-regulating *causal nexi*. God was merely the "First Cause", and could intervene in the world only by breaking or suspending the "natural laws". Locke and Hume used the determinism of newtonian physics to argue that natural laws were inviolable, and therefore that miracles could not happen⁹. Their conclusion seemed to be vindicated in the nineteenth century when the dar-

winian revolution purged from biological systems the simple notion of purpose and created pattern. And as Cupitt says, "religion was more badly shaken when the universe went historical in the nineteenth century than it had been when it went mechanical in the seventeenth century"¹⁰. The futility of believing in a god unable to do anything exposed the problem that spurred the English bishops to re-affirm that miracles could happen¹¹.

Miracles and mechanisms

Defenders of miracles have tended to descend into an unconvincing mysticism or an assault on determinism. A few decades ago, it was fashionable to claim that physical indeterminacy gave God enough freedom to control events. Biological indeterminacy is a live debate now, particularly in sociobiology¹². For example, Lewontin (unlikely to argue that miracles are common or important) strongly attacked the reality of biological laws beyond "very special rules of comportment or particular physical entities If we are to find biological laws that can be the models for social laws, they will surely be at the level of laws of population, laws of evolution, laws of organization. But is it precisely such laws that are absent in biology, although many attempts have been made to erect them"¹³.

However, the case for miracles does not depend on indeterminacy, since the intellectual orthodoxy stemming from Hume's underlying thesis is not as strong as it is usually made out to be. C.S. Lewis pointed this out succinctly: "We must agree with Hume that if there is absolutely 'uniform experience' against miracles, if in other words they have never happened, why then they never have. Unfortunately we know the experience against them to be uniform only if we know that all the reports of them are false. And we know the reports to be false only if we know already that miracles have never occurred. In fact, we are arguing in a circle"¹⁴.

Exposing the fallacy of Hume's attack on miracles also reveals that it is based on an unjustified assumption, that events have only a single cause and can be fully explained if that cause is known. This is logically wrong. For example, an oil painting can be "explained" in terms either of the distribution of pigments or the intention and design of the artist; both explanations refer to the same physical object but they complement rather than conflict. In

the same way, a miracle may be the work of (say) a divine up-holder of the physical world rather than a false observation or unknown cause. Such an interpretation does not depend on any irruption into a causal network, since the determinism of the machine is only one of the levels of the phenomenon (*sensu* Polanyi¹⁵).

"Complementary" explanations of causation are excluded only by making the reductionist assumption that a single identifiable cause is the sole effect operating in a particular situation. This assumption is common, but unnecessary and restrictive. Medawar has dissected this clearly: "That there is indeed a limit upon science is made very likely by the existence of questions that science cannot answer and that no conceivable advances of science would empower it to answer. These are the questions that children ask — the 'ultimate questions' of Karl Popper. I have in mind such questions as: How did everything begin? What are we all here for? What is the point of living? Doctrinaire positivism — now something of a period piece — dismissed all such questions as nonquestions or pseudoquestions such as only simpletons ask and only charlatans of one kind or another profess to be able to answer. This peremptory dismissal leaves one empty and dissatisfied because the questions make sense to those who ask them, and the answers to those who try to give them; but whatever else may be in dispute, it would be universally agreed that it is not to science that we should look for answers. There is then a *prima facie* case for the existence of a limit to scientific understanding."¹⁶

As far as miracles are concerned, this means that they are impossible to prove or disprove on normal scientific criteria; we accept the possibility of their occurrence by faith, and equally deny them by faith. Even if we know or deduce the mechanism behind a miracle, this does not necessarily remove the miraculous element. For example, the Bible tells us that the Israelites crossed the Red Sea dry-shod because "the Lord drove the sea back by a strong east wind all night and made the sea dry land" (*Exodus* 14:21); the significance of the miracle lies in its timing and place rather than its actual occurrence.

Implications

The act of faith that denies the possibility of miracles is a straightforward reductionist judgement. Miracles by themselves are always susceptible to an explanation other than the miraculous (even if they have physical manifestations, such as "spontaneous" healing or the Empty Tomb), so the value of the reductionist assumption can be best tested by its implications. These were spelt out with depressing clarity in the nihilism of Jacques Monod¹⁷, and comprehensively answered by W.H. Thorpe¹⁸ who expounded a version of the

dualism of Sherrington, Eccles and Popper which is kin to the complementarity espoused above¹⁹.

There are implications of embracing a reductionist determinism which impinge on two recent controversies: creationism and the definition of human life. Creationism is largely an insistence that God made the world in a particular way, without using "normal" evolutionary mechanisms. Part of this claim stems from a restricted interpretation of the Bible, but it has the effect of prescribing that God acted in an interventionist fashion. Notwithstanding, it is entirely consistent with both evolutionary biology and Bible texts to maintain that God worked "complementarily" with genetic processes so that the world is both a causal outcome of mutation, selection, and so on, but *also* a divine creation. The creationist position is at odds with both scientific and theistic understanding^{20,21}. Individual human life has a physiological and genetic continuity with that of other humans (and indeed, other animals); the *value* of individual life lies not in genetic uniqueness (cancers and hydatidiform moles are also genetically unique) but in being (in Christian language) "made in the image of God". This *imago* is not a physical entity, and it is a category mistake to confuse it with genetic coding or mental function. Notwithstanding, defenders of the inviolability of the early embryo make this precise mistake. The *imago* is a non-biological attribute, and there is no logical (or scriptural²²) reason for assuming that it is present from conception. If this simple point was realized, the ethical debate over developments in human reproduction could proceed more sensibly.

The conventional view of miracles is that they depend on supernatural intervention in, or suspension of, the natural order. Some theologians have been over-impressed with scientific determinism, and have attempted a demythologized (miracle-free) religion. This endeavour is now unfashionable, but it is worse than that; Nebelick called it "a speculative-device imposed on unsuspecting persons... based on false presuppositions about both science and the scientific world view"²³. This is no help to scientists, and an interventionist God will always be an embarrassment to us.

I believe that the interpretation that miracles are a necessary but unpredictable consequence of a God who holds the world in being is more plausible and more scriptural than deist interventionism. This does not mean that apparent miracles should be approached with any less objectivity than we would employ for any scientific observation; our standards of evidence should be just as rigorous. Those who deny the possibility of miracles are exercising their own brand of faith; this is based on a questionable assumption, and

one which creates problems with its implications. Miracles in the New Testament are described as unusual events which are wonders due to God's power, intended as signs. Confining oneself wholly to this category (leaving aside the question of whether other sorts of miracles occur), this makes at least some miracles expectable and non-capricious, and independent of any knowledge of their mechanism.

In his exposition of the "Two Cultures", C.P. Snow described the scorn of the one for the other as intellectual Ludditism²⁴. Miracles are examples of events which may easily be denied by an illegitimate reductionist Ludditism; scientific reality will be hindered in the process. A doctrinaire disbelief in miracles is not "more scientific" than a willingness to accept that they may occur. Some years ago Sir George Porter wrote: "Most of our anxieties, problems and unhappiness today stem from a lack of purpose which was rare a century ago and which can fairly be blamed on the consequences of scientific inquiry. . . . There is one great purpose for man and for us today, and that is to try to discover man's purpose by every means in our power. That is the ultimate relevance of science"²⁵. He was not writing specifically about miracles, but his argument applies. Miracles are not inherently impossible or unbelievable, and acceptance of their existence does not necessarily involve credulity, but does involve recognizing that science has limits. □

Professor R.J. Berry is at the Department of Zoology, University College London, Gower Street, London WC1E 6BT, UK.

1. *The Times* (London) (13 July 1984).
2. *Nature* **310**, 171 (1984).
3. Clarke, P.G.H. *Nature* **311**, 502 (1984).
4. MacKay, D.M. *Nature* **311**, 502 (1984).
5. *The Nature of Christian Belief* (Church House, London, 1986).
6. *The Times* (London) (6 June 1986).
7. Polkinghorne, J. *The Way the World Is* (SPCK, London, 1983).
8. Coulson, C.A. *Science and Christian Belief* (Oxford University Press, 1955).
9. Brown, C. *Miracles and the Critical Mind* (Eerdmans, Grand Rapids, 1984).
10. Cupitt, D. *The Sea of Faith* (BBC, London, 1984).
11. Harris, M.J. *Easter in Durham* (Paternoster, Exeter, 1985).
12. Berry, R.J. *Free to be Different* (ed. Stott, J.R.W.) Ch.5 (Marshall, Morgan & Scott, Basingstoke, 1984).
13. Lewontin, R.C. *Population and Biology* (ed. Keyfitz, N.) Ch.1 (Ordina, Liege, 1985).
14. Lewis, C.S. *Miracles* (Bles, London, 1974).
15. Polanyi, M. *Knowing and Being* (Routledge & Kegan-Paul, London, 1969).
16. Medawar, P. *The Limits of Science* (Harper & Row, New York, 1984).
17. Monod, J. *Chance and Necessity* (Collins, London, 1971).
18. Thorpe, W.H. *Purpose in a World of Chance* (Oxford University Press, 1978).
19. MacKay, D.M. *Freedom of Action in a Mechanistic Universe* (Cambridge University Press, 1979); *Human Science and Human Dignity* (Hodder & Stoughton, London, 1979).
20. Midgley, M. *Evolution as a Religion* (Methuen, London, 1985).
21. Berry, R.J. *Epworth Rev.* **13**, 74–85 (1986).
22. Rogerson, J.W. *Abortion and the Sanctity of Human Life*, (ed. Channer, J.H.) Ch.4 (Paternoster, Exeter, 1985).
23. Nebelick, H.R. *Scot. J. Theol.* **37**, 239 (1984).
24. Snow, C.P. *The Two Cultures* (Cambridge University Press, 1975).
25. Porter, G. *The Times* (London) (21 June 1975).

Terrestrial origin of the Moon

A. E. Ringwood

Research School of Earth Sciences, Australian National University, Canberra, ACT 2601, Australia

Abundance patterns of siderophile elements in the Earth's mantle have been determined by complex processes connected with internal segregation of a metallic core amounting to 32% of the Earth's mass. Although the lunar core may be only 2% of the lunar mass and formed in very different conditions to the Earth's core, the lunar siderophile pattern is closely related to that of the Earth's mantle, implying that the material in the Moon derived mainly from the Earth's mantle after core formation. New models for the accretion of the Earth which require impacts by large planetesimals in the presence of a corotating primitive terrestrial atmosphere provide a mechanism for transferring material from the Earth's mantle into geocentric orbit as a ring of proto-lunar planetesimals, from which the Moon later accreted.

MORE than a century ago, Darwin¹ studied the tidal evolution of the Earth-Moon system and concluded that the Moon had been torn out of the Earth's mantle by a tidally induced fission process. Subsequently, it was realized that this would explain the anomalously low density of the Moon compared with the bulk Earth. Darwin's hypothesis enjoyed several decades of popularity before lapsing into obscurity after Jeffreys² had demonstrated that the proposed fission mechanism was physically unsound. Thereafter, it was generally considered that the low density of the Moon could be explained by an extension of the same processes responsible for causing marked differences between the intrinsic densities of the terrestrial planets (see refs 3, 4). Essentially, the Moon was regarded as having formed by processes analogous to those experienced by independent planets.

Jeffreys³, Urey⁴ and many others explained the intrinsic density variations between the terrestrial planets by assuming that they are composed of varying proportions of silicate (initial density, $\rho_0 \approx 3.3 \text{ g cm}^{-3}$) and nickel-iron ($\rho_0 = 7.9 \text{ g cm}^{-3}$) phases, each being essentially of constant composition. According to this model, Mercury and Mars contain ~65% and 20% of metal phase respectively whilst Earth and Venus possess intermediate metal contents. The model implies that the Moon contains <5% of metal phase. Urey^{5,6} proposed that physical processes occurring in the solar nebula before accretion of the planets had caused variable degrees of fractionation of iron particles from silicate particles in different regions of the nebula. Subsequent accretion of planets in particular regions reflected these pre-existing metal/silicate inhomogeneities. The major role attributed to metal/silicate fractionation in the solar nebula was supported by measurements which seemed to show that the abundance of iron in the Sun was lower by a factor of 5 than is found in Earth, Mars and chondritic meteorites. Urey's interpretation became widely accepted and has been reflected in subsequent cosmogonic models^{7,8}.

Since 1959, I have argued that the mechanisms invoked by the above hypotheses to explain the metal/silicate fractionations were physically implausible and therefore attempted to develop a cosmogonic model aimed at minimizing the role of this process⁹⁻¹¹. The model proposed that the relative abundances of iron and the common lithophile elements (Mg, Si, Ca, Al) were the same in Mars, Earth and Venus, and were similar to those in the Sun and in chondritic meteorites. Differences in density between Mars, Venus and Earth were caused by differing oxidation states, a variable which was readily explained on cosmochemical grounds.

This interpretation received substantial support in 1969 when a more accurate determination of the relative abundance of iron in the Sun showed that it was similar to that in chondrites¹². Moreover, new space probe measurements have since provided precise values of the densities of Venus and Mars and of the martian moment of inertia. It was readily demonstrated that the gross physical (and chemical) properties of Venus, Earth and Mars were explicable in terms of models based on chondritic abundances of major elements, differing in mean oxidation state, with Mars being significantly more oxidized than Earth and Venus¹¹. The only planet displaying the effects of major metal/silicate fractionation is Mercury. Several authors have suggested that its high density, implying a high iron content, may be caused by specialized conditions of accretion, owing to its position nearest to the Sun^{8,10}.

These considerations have an important bearing on the origin of the Moon. During the period 1950-70, most scientists treated the Moon as representing an extreme case of the general process of iron/silicate fractionation which was believed to have occurred between Sun and planets, and between the planets themselves. In this sense, it was regarded as an 'independent planet', which, owing to special circumstances, had accreted in orbit around the Earth or had been captured into Earth orbit.

However, we now recognize that the Moon must have accreted in a region of the Solar System between Earth, Mars and Venus, and probably close to the Earth. Currently, there is no evidence requiring iron/silicate fractionation between these planets and the Sun. From this perspective, the high depletion of metallic iron in the Moon must be recognized as a truly remarkable phenomenon. This led me in 1960 to revive Darwin's hypothesis that the material now in the Moon was derived from the Earth's mantle by a fission process, albeit employing a different mechanism¹³ from that proposed by Darwin. Analogous fission-related hypotheses were also developed during the 1960s by Wise¹⁴, Cameron¹⁵ and O'Keefe¹⁶. Subsequently, I abandoned the fission mechanism and developed another model in which the Moon was formed from material evaporated from the Earth's mantle in a high-temperature regime caused by rapid accretion of the Earth^{10,11,17-19}.

If the Moon were indeed derived from the Earth's mantle, one might hope to find supporting evidence in a comparison of the chemical compositions of both bodies. Unfortunately, there are divergent views among geochemists regarding the major element composition of the Moon. Some believe that the Moon's composition is similar to that of the Earth's upper mantle^{11,20,21}, whereas others^{22,23} believe that refractory elements are markedly

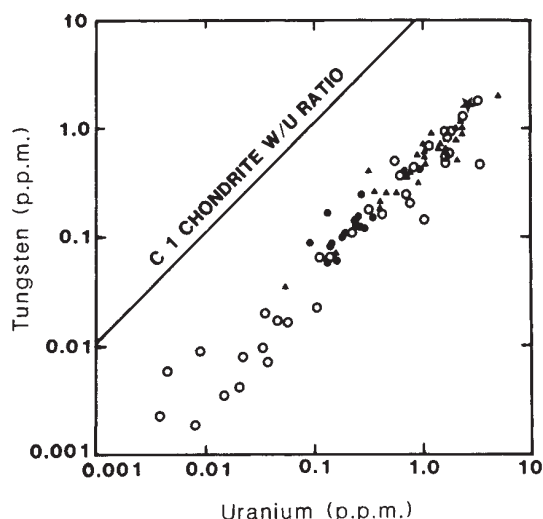


Fig. 1 Tungsten (siderophile) and uranium (lithophile) concentrations in terrestrial and lunar rocks. $\circ$, Earth's mantle; $\star$, upper crust; $\bullet$, lunar mare; $\blacktriangle$, lunar highlands. The W/U ratios in both bodies are remarkably similar despite the fact that core formation within the Earth caused tungsten to be relatively depleted in the upper mantle by a factor of ~ 25 compared with the primordial C1 W/U ratio (solid line). The lunar samples include both mare and highland rocks, showing that the lunar W/U ratio is of global significance. (After refs 33, 49, 52.)

enriched in the Moon. But is there any other source of compositional evidence which points unambiguously towards a terrestrial origin? I believe that the answer is provided by the siderophile geochemistry of Earth and Moon.

Siderophile geochemistry

The abundance patterns of siderophile elements in the Earth's mantle display some remarkable characteristics. For example, Ni, Co, Cu, Fe, Ga, W, Mo, P, As, Sn, Ag and Ge are present at levels ranging between 1 and 16% of their primordial abundances²⁴⁻²⁸. Little relationship exists between the partition coefficients for these siderophile elements between silicate and metallic iron phases, and their relative abundances in the mantle. The platinumoids, Au, Re, S, Se and Te, are present in the mantle at $\sim 0.3\%$ of their primordial abundances^{25,27-29}. Most of the above siderophiles are considerably overabundant in the mantle compared with expectations based on experimentally determined partitions between silicate and metallic iron phases.

Chromium and vanadium, which display distinct siderophile tendencies above 1,500 °C (refs 30, 31), are significantly depleted in the mantle by factors of 2 and 1.5, respectively²⁵. These depletions almost certainly reflect their partial entry into the core³². Manganese also displays a fourfold depletion which may be caused either by entry into the core^{20,32,33} and/or by loss due to volatility²⁴.

The above patterns have evidently been caused by a combination of several complex processes, as yet incompletely understood, which occurred during the segregation of a metallic core, amounting to 32% of the Earth's mass. Partitions of siderophiles would have been affected by the very high pressure (P) and temperature (T) regime under which this process operated, and by the presence in the core of $\sim 10\%$ of light elements, probably consisting mainly of oxygen³⁴. It is also widely believed that the terrestrial siderophile pattern was substantially influenced at an advanced stage of accretion by the physical mixing into the mantle of an oxidized, solar nebular condensate in conditions which did not permit equilibration with metal phase^{20,27-29}.

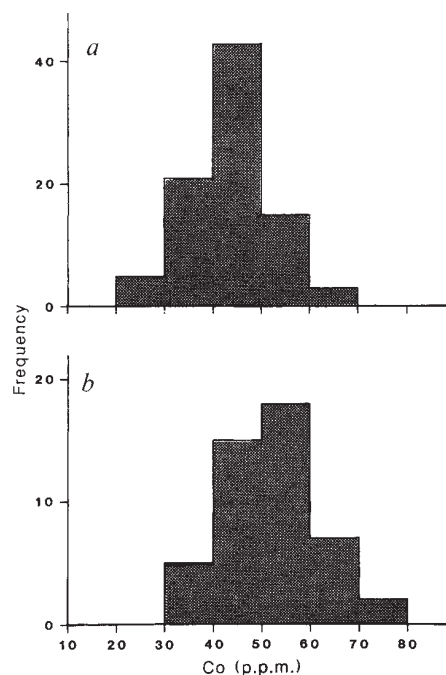


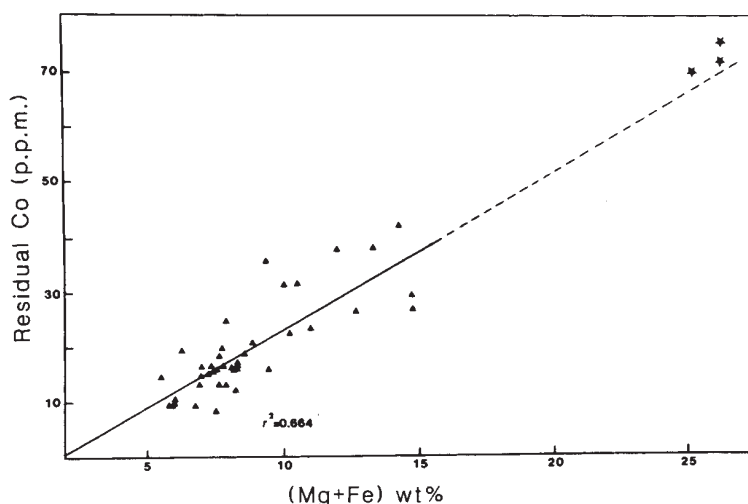
Fig. 2 Comparison of Co abundances in terrestrial oceanic tholeiites (*a*, mean Co = 43 p.p.m.) and in lunar low-Ti mare basalts (*b*, mean Co = 51 p.p.m.) (from ref. 53).

The particular combination of processes which generated the siderophile signature of the Earth's mantle probably did not operate to the same extent in all planets and differentiated planetesimals. Accordingly, it is expected to be unique to the Earth, or to an Earth-like planet. This expectation can be tested by comparing the terrestrial siderophile signature with the siderophile signature of the parent body of shergottites³⁵⁻³⁸ which is widely identified as Mars^{39,40}. As with the Earth, most siderophiles (such as, Co, Ni, P, W, Ge, Mo, Au and Ir) are substantially overabundant compared with expectations based on equilibrium partitions between silicate and metallic iron phases. There are, however, very significant differences between the martian and the terrestrial signatures. In shergottites, W and P are present at levels about 5-10 times higher than the Earth, while Cu, Ni, As, Tl and In are depleted by factors of 5-20. Moreover, Mn and Cr are present in primordial C1 abundances, whereas these elements are depleted in the Earth. In addition, the oxygen isotope composition of shergottites is significantly different from that of the Earth^{41,42}. Dreibus and Wänke³⁶ have pointed out that the differences between the martian and terrestrial siderophile signatures could be explained by differences in the mode of accretion of both planets, combined with the presence of a large quantity of sulphur in the martian core.

The siderophile signature displayed by the silicate phase of the eucrite parent body (EPB) differs in turn from those of the Earth and Mars. In this case, the siderophile abundances reflect a close approach to chemical equilibrium between silicate and iron phases during segregation of the core in the EPB. In consequence, many siderophiles (for example, Ni, Co, Cu, Ga, Mo and Ir) are relatively depleted by factors varying between 3 and 50 in the EPB silicate phase compared with the Earth's mantle⁴³⁻⁴⁵. On the other hand, Mn, V and Cr are present in primordial abundances, in contrast to their depletion in the Earth's mantle^{32,43}.

The above discussion highlights the unique nature of the siderophile signatures in the mantles of planets and planetesimals which have differentiated to produce substantial metallic cores. The siderophile signature of the Moon, in which

Fig. 3 Cobalt in primitive lunar volcanic glasses (★) and Apollo 16 highland breccias (▲, corrected for meteoritic contamination) plotted against (Mg+Fe). Note that the solid line representing the best fit to the highland breccia data projects directly into the field of primitive volcanic glasses, showing that the lunar Co/(Mg+Fe) ratio is of global significance. (After ref. 53.)



a metallic core probably amounts to <2% of its mass, is therefore likely to be of considerable interest.

Lunar siderophile signature

The importance of lunar siderophile geochemistry to the problem of the Moon's origin was emphasized by O'Keefe⁴⁶ and O'Keefe and Urey⁴⁷. A comparative study of siderophile abundances in terrestrial and low-Ti lunar basalts, carried out by Ringwood and Kesson²⁴, concluded that the average abundances of a particular group, including Fe, Co, Ni, W, P, S and Se, were similar within a factor of ~2. Pointing to the unique factors responsible for the siderophile signature of the terrestrial mantle and its derived basalts, they concluded that "the similarity in siderophile elements between the Moon and Earth's mantle therefore implies that the Moon was derived from the Earth's mantle after the Earth's core had segregated". They also recognized that another group of siderophiles (Ga, Cu, Ge, As, Ag, Au, Sb, Re) was genuinely depleted in the Moon compared with Earth and ascribed this to loss as volatile species. Delano and Ringwood⁴⁴ made a parallel study of the siderophile chemistry of lunar highland breccias, after applying corrections for meteoritic contamination. They found that the indigenous siderophile signature in the lunar highlands closely resembled that of the low-Ti mare basalts and terrestrial basalts, thereby reinforcing the previous conclusion.

Wänke and co-workers in Mainz independently arrived at very similar conclusions^{48,49}, based on detailed studies of the geochemical behaviour of W and P in terrestrial rocks and in lunar highland rocks and basalts (Fig. 1). Wänke *et al.*^{50,51} extended these conclusions by demonstrating the occurrence of a substantial component in the Apollo 16 highlands breccias, possessing very similar Ni/Mg and Co/Mg ratios to the Earth's mantle. Wänke and Dreibus^{20,32,33} also drew attention to the genetic significance of similar Cr, V and Mn abundances in the Moon and in the Earth's mantle. As noted earlier, these elements are significantly depleted in the mantle, primarily because of their partial entry into the Earth's core. Wänke *et al.*⁵⁰ also recognized that another group of highly siderophilic elements was depleted in the Moon compared with the Earth's mantle. They attributed this depletion not to volatility (as proposed by Ringwood and Kesson²⁴) but to selective extraction by segregation of a small metallic core within the Moon.

Tungsten and cobalt provide good examples of moderately siderophilic elements whose geochemical behaviour in the Earth are well understood. In oxidized form, W is an incompatible element and behaves similarly to U during magmatic fractionation. Hence the W/U ratio in terrestrial rocks remains approximately constant over a wide concentration range (Fig. 1)^{33,49,52}

Tungsten and uranium are both refractory elements and were accreted by the Earth in chondritic relative abundances^{23,49}. Because of its siderophile nature, W was depleted in Earth's mantle by a factor of 25 owing to its entry into the Earth's massive core. The W/U ratios of lunar mare and highland rocks are almost identical to those of terrestrial rocks (Fig. 1). In view of the similar U abundances in the Earth's mantle and Moon¹¹, the W abundances in both bodies must display a corresponding similarity.

Unlike W, Co is a compatible element which is not strongly fractionated (relative to Mg and Fe) during magmatic differentiation. Abundances of Co in terrestrial tholeiites and lunar low-Ti mare basalts are shown in Fig. 2 and are very similar^{24,53}. Figure 3 demonstrates a good correlation between Co and (Mg+Fe) in lunar highland breccias. Moreover, the highland Co/(Mg+Fe) ratio is identical to that displayed by the least fractionated lunar mare basaltic magmas as represented by volcanic glasses, confirming that the lunar Co/(Mg+Fe) ratio is also a global characteristic. Experimental investigation⁵³ of the distribution of Co between a primitive lunar volcanic glass, liquidus olivine and metal phase confirmed that the Co content of the lunar mantle was very similar to that of Earth's mantle.

The conclusion by both Canberra and Mainz groups that the similarity in abundances of a particular group of siderophile elements in the Earth's mantle and Moon unequivocally implies that the material in the Moon has been derived from the Earth's mantle created considerably controversy. Most lunar scientists were reluctant to accept the proposition that the vexed question of the Moon's origin could be settled by such a 'simple' argument. Nevertheless, the controversy had beneficial results because it stimulated the acquisition of additional analytical data on the abundances of siderophile elements in lunar and terrestrial rocks, and on partition coefficients of siderophiles between metal and silicate phases. The demonstration⁵² that molybdenum was depleted in the Moon by a factor of ~30 was particularly significant, since the depletion of this highly siderophilic element could not be ascribed to volatility, and pointed towards extraction by a small metallic lunar core⁵².

The new data have been combined with earlier results in Fig. 4 as a plot of abundance ratios of siderophile elements in the Earth's¹¹ mantle and Moon versus their metal/silicate partition coefficients. Analogous but less comprehensive diagrams have been constructed by Wänke and Dreibus³³ and by Newsom⁵². It is seen that the abundances of elements less siderophilic than nickel (Mn, V, Cr, Fe, W, Co, P, S, Se) are similar (within a factor of 2) in both bodies. Elements more siderophilic than nickel (Cu, Mo, Re, Au) are depleted to degrees which correspond well with their metal/silicate partition coefficients. Nickel

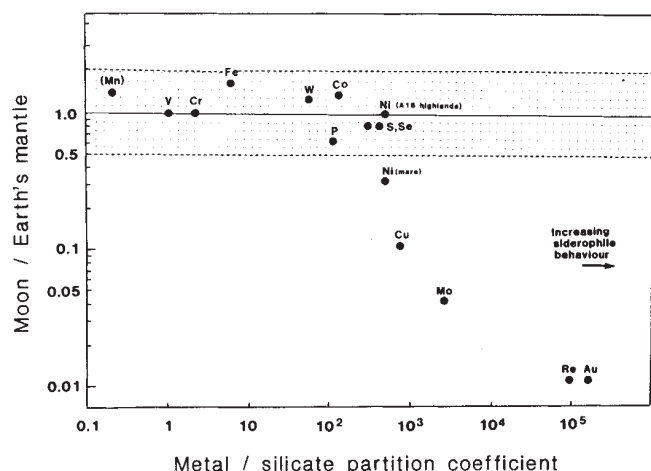


Fig. 4 Abundance ratios of siderophile elements in the Moon and Earth's mantle versus their metal/silicate partition coefficients. Partition coefficients used for compatible elements, Mn, V, Cr, Fe, Co and Ni, are based on measured distributions between olivine and Fe-metal phases^{30,31,53,79}. Partition coefficients for generally incompatible elements W, P, S, Mo, Re and Au, are for basaltic melts and Fe-metal phases^{49,80-83}. Positions of Se and Cu were estimated by interpolation from the free energies of formation of their oxides. Abundance data for siderophile elements in the Moon and Earth's mantle are from the following sources: Mn, refs 24, 25; V, Cr, refs 24, 25, 79; Fe, refs 24, 25; W, refs 24, 25, 49, 52; P, refs 24, 25, 33, 44; Co, refs 24, 25, 53; S, Se, ref. 24; Ni, refs 50, 53; Cu, refs 24, 25; Mo, refs 27, 52; Re, Au, refs 24, 29, 84, 85. Brackets around Mn indicate the possibility that its depletion in the Moon and Earth might be related to volatility rather than siderophile behaviour²⁴.

presents an interesting case and has been studied extensively; it is depleted by a factor of ~ 3 in the source regions of the most primitive lunar basalts^{33,53-55} but is present in terrestrial abundances (relative to Mg) in breccias at the A16 lunar highlands site^{50,53}.

Wänke and colleagues^{33,50,52} have shown that the depletions (below terrestrial levels) of highly siderophilic elements shown in Fig. 4 can readily be explained by separation of a small metallic core amounting to $<1\%$ of the lunar mass. This would not markedly affect the abundances of moderately siderophilic elements but would have strongly depleted the highly siderophilic elements. The presence of a small ($<2\%$ by mass) lunar core is strongly suggested by the lunar moment of inertia coefficient⁵⁶ ($I/MR^2 = 0.391 \pm 0.002$) and by the observation of a phase shift in the forced precession of the lunar figure⁵⁷.

The composition of the lunar core is constrained by the nickel and cobalt abundances of lunar mantle silicates. The nickel content of the lunar mantle can be determined using a procedure identical to that used to determine its abundance in the Earth's mantle, that is by measuring the nickel content in the olivine crystallizing on the liquidus of the most primitive, unfractionated basaltic magmas derived from the lunar interior. Ringwood and Seifert⁵³ used this technique to demonstrate that the olivine of the lunar mantle contains ~ 900 p.p.m. Ni. Moreover, they showed that any metal in equilibrium with lunar olivine would contain at least 40% nickel. It is reasonable to expect that lunar mantle silicates would have equilibrated with metal phase during segregation of the lunar core. Thus the lunar core is required to contain at least 40% Ni in sharp contrast to the Earth's core, which contains $\sim 5\%$ Ni (ref. 53). An analogous technique has been applied⁵³ to a totally independent data-set to demonstrate that the cobalt content of the lunar core is also higher than that of the Earth's core. The gross dissimilarities in chemical compo-

sitions between the lunar and terrestrial cores have important implications for hypotheses of lunar origin.

Implications for lunar origin

Previously, it was shown that the lunar and terrestrial siderophile signatures differ fundamentally from those of the shergottite parent body (Mars) and the EPB. The reasons why the terrestrial siderophile signature is likely to be unique to the Earth or an Earth-like planet have also been documented. Figure 4 provides a striking illustration of the close relationship between the abundances of siderophiles in the Earth's mantle and the Moon. The moderately siderophilic elements in the Moon are clearly 'Earth-like'. Although the highly siderophilic elements are greatly depleted in the Moon compared with Earth, this can be readily explained by segregation of a small amount ($<1\%$) of a Ni-rich metal phase ($\sim 40\%$ Ni) from parent material of terrestrial composition to form a lunar core. I conclude that the close relationship between the lunar and terrestrial siderophile signatures depicted in Fig. 4 provides definitive evidence for the terrestrial origin of the Moon, as argued previously by the Canberra and Mainz groups. This is further supported by the identity in oxygen isotope compositions between Earth and Moon and their dissimilarity with those of shergottites and differentiated meteorites^{41,42}.

Could these abundance patterns be explained in terms of an origin of the Moon as an "independent planet"?²³ Newsom⁵² has attempted to show that the lunar abundances of W, P, Ni and Co could be explained by this hypothesis. His model requires the separation within the Moon of an iron-rich metallic core amounting to $\sim 5\%$ of the lunar mass. The model can be rejected on the following grounds.

- (1) Current geophysical and petrological models of the Moon agree that its FeO content lies between 8 and 17% (refs 21, 50). If the lower limit is accepted, an iron core amounting to 5% of the lunar mass would lead⁵⁸ to a mean lunar density exceeding 3.40 g cm^{-3} in contradiction to its observed density of 3.34 g cm^{-3} . The discrepancy is increased if the more widely accepted estimate of 13–16% FeO in the lunar mantle is used.
- (2) Experimental data cited previously show that the lunar core is Ni-rich; this causes a decrease in the effective metal/silicate partition coefficients for W (ref. 49), P (ref. 59) and Co (ref. 53) as compared with pure iron. Newsom's model therefore requires a metal core constituting substantially more than 5% of the lunar mass.
- (3) The size of the lunar core is constrained to be $<2\%$ of the lunar mass by the seismic observations of Nakamura *et al.*⁶⁰ who state that "the radius of a low-velocity (molten) core cannot be greater than about 350 km because normal transmission of P waves to that depth is observed".
- (4) The model fails to explain the similar abundances of Mn, Cr, V, S and Se in the Moon and the terrestrial mantle (Fig. 4) and their variable depletions as compared with primordial abundances.

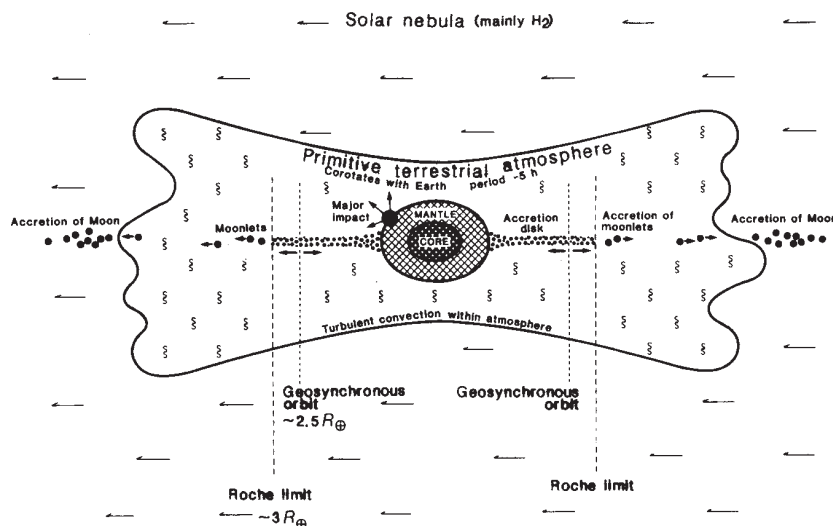
More generally, the 'independent planet' hypothesis of lunar origin encounters a series of additional difficulties, which cumulatively are quite lethal.

(i) The siderophile signature of the terrestrial mantle was established by a series of complex processes occurring during the segregation of a large ($32\% M_E$) metallic core within the Earth. If the Moon had formed as an independent planet, it would be quite incredible if its very different conditions of accretion, followed by segregation of a relatively small ($\sim 2\% M_m$) core possessing a composition grossly different to the Earth's core, and formed in a totally different P , T environment, had, nevertheless, produced a siderophile signature similar to the Earth's mantle.

(ii) The dilemma of providing a plausible mechanism for explaining the depletion of iron in the Moon, as compared with Earth, Venus and Mars, is not resolved.

(iii) Previous attempts to solve this problem have relied on

Fig. 5 Model showing the formation of the Moon via the ejection of material from the Earth's mantle by impacts from late-accreting planetesimals in the presence of a corotating primitive terrestrial atmosphere.



highly speculative physical mechanisms to fractionate metal from silicate in the nebula, before accretion of the Moon⁶¹. However, in these models, the composition of the metal phases accreted by Earth and Moon would be similar, that is Fe-rich and Ni, Co-poor. This is contrary to the evidence cited previously that the lunar core is actually Ni, Co-rich and quite different in composition to the Earth's core.

(iv) Some workers⁶² have attempted to interpret the Moon as having formed from an earlier generation of planetesimals which accreted in the solar nebula and then melted and differentiated analogously to the eucrite parent body. These were subsequently fragmented and their metal cores somehow removed, leaving the Moon to accrete from the differentiated silicate mantles. The major differences already noted between the siderophile signatures of the Moon and the silicate phase of the EPB are unexplained by this model. Moreover, the mechanisms proposed for removing the metal cores are highly implausible^{63,64}.

Making of the Moon

The geochemical evidence reviewed previously can apparently only be satisfied if the Moon was formed from material derived from the Earth's mantle after the core had segregated. Provision of the energy and angular momentum required to remove >1% of the Earth's mass and place it into orbit represents a rather formidable challenge. For two decades, I have argued that the energy was provided by the Earth's own gravitational energy of accretion and that the requisite angular momentum was transferred from the rapidly rotating Earth to protolunar material by a corotating primitive terrestrial atmosphere.

According to these models^{10,11,17-19}, accretion of the Earth occurred before dissipation of the primordial gases of the solar nebula. Accretion was accompanied by the formation of a primitive terrestrial atmosphere comprised mainly of hydrogen gravitationally captured from the solar nebula⁶⁵⁻⁶⁷. The primitive atmosphere was coupled to the Earth's rotation through turbulent viscosity and hydromagnetic torques and was thereby spun out into a corotating disk (period ~5 h). During the later stages of accretion of the Earth, high temperatures were produced by a combination of rapid accretion and thermal insulation by the primitive atmosphere. In these conditions, material from the mantle was evaporated into the primitive atmosphere and spun out into the disk. As the primitive atmosphere cooled and was dissipated, the silicate components were precipitated to form a ring of Earth-orbiting planetesimals. Further fractionation due to volatility occurred during the precipitation process, since the more volatile components would be precipitated at relatively low temperatures, forming micrometre-sized smoke

particles. These remained coupled to the escaping gases and hence were removed from the system. The Moon accreted from the ring of devolatilized Earth-orbiting planetesimals.

The principal problem with this scenario was the short timescale (~10⁶ yr) over which the Earth must have formed if the gravitational potential energy of impacting planetesimals was to have generated the sustained high temperatures necessary to evaporate part of the upper mantle during the late stages of accretion. An attractive alternative is provided by the models of Hartmann and Davis⁶⁸, and of Cameron and Ward⁶⁹. These authors invoke one or a few impacts by giant, late-accreting planetesimals to evaporate material from the Earth's mantle and place it into orbit. The planetesimals are believed to have sizes ranging between a substantial fraction of the lunar mass and the size of Mars⁷⁰. The physical processes envisaged⁷¹ are quite closely related to those of my earlier models. Rather than achieve high mantle-evaporation temperatures via the 'steady-state' liberation of gravitational potential energy from a continuum of relatively small impacts over a short accretion timescale (1 Myr), giant impact models achieve these conditions in one or a few transient ultrahigh energy events, and permit a much longer accretion timescale for the Earth (such as 10⁷-10⁸ yr).

Although the giant impact model is currently attracting considerable interest, it is not free from potential problems. It is expected to lead to complete melting of the entire Earth^{70,72}, this should have caused efficient differentiation of the mantle accompanied by formation of a massive 4,500-Myr terrestrial crust, contrary to observation¹¹. Complete melting of the Earth and subsequent crystallization of the mantle should also have been accompanied by near-quantitative degassing of volatiles such as CO₂, N₂, Cl and inert gases. This implication is not readily reconciled with available geochemical evidence^{4,5,73}.

Accordingly, I now prefer a model intermediate between my earlier rapid-accretion models and the current giant impact models (Fig. 5). A key role in lunar genesis continues to be played by the primitive terrestrial atmosphere which corotates with the Earth, extending as a disk perhaps out to 5 Earth radii (R_⊕)^{65,74,75}. The atmosphere would extend the effective radius of capture of incoming planetesimals. This factor might have contributed to a relatively short rotation period for the Earth.

During most of the Earth's accretion, small planetesimals would be captured by the atmosphere and would spiral down to the equatorial plane, forming an accretion disk. Large planetesimals would impact the Earth directly; their ejecta would be captured by the atmosphere and transferred to the accretion disk. Within the disk, planetesimals orbiting inside the geosynchronous limit would experience gas drag and spiral

inwards to the Earth. Beyond the geosynchronous radius, the atmosphere, coupled to the Earth by hydromagnetic torques, would be rotating faster than the keplerian velocities of the planetesimals, which consequently would be driven outwards. However, as the mass of the Earth continued to grow by accretion, the geosynchronous orbit expanded so that most of this material was eventually accreted by the Earth.

It is possible to form the Moon only at a late stage of accretion when the Earth had developed nearly to its present size and core formation was complete (or nearly complete). At this stage, impacts of many large ($100 < R < 1,000$ km) but not necessarily giant ($R > 1,000$ km) planetesimals at velocities exceeding 12 km s^{-1} would vapourize ~ 5 times their own mass of target material and would shock-melt 100 times their own mass⁷⁶. The shock-melted material would probably form a spray of devolatilized droplets. Rapid expansion of the impact cloud would cause acceleration to high velocities^{70,71}. As the cloud expanded and cooled, selective condensation of the less volatile components would have produced more liquid droplets which subsequently solidified. Highly volatile elements condensed only at relatively low temperatures, forming smoke particles.

The corotating primitive atmosphere is believed to have had an essential role in transferring angular momentum from the solid Earth to impact-evaporated gases and liquid spray, so that a substantial proportion of the ejected material was placed in circular, equatorial orbits. The devolatilized droplets (1–10 mm) and coarser condensates accreted to produce an assemblage of devolatilized planetesimals. Most planetesimals probably

accreted in orbits smaller than the geosynchronous limit ($\sim 2.5 R_{\oplus}$ for a rotation period of 5 h). They would have lost energy through dissipation in the corotating atmosphere and hence spiralled back to Earth. However, a significant proportion of planetesimals would have been formed beyond the geosynchronous limit. They would have been accelerated by gas-drag, causing them to spiral outwards, beyond Roche's Limit ($\sim 3 R_{\oplus}$). Here they would have accreted into moonlets, which continued to spiral outwards (through gas drag) to the boundary region between the corotating primitive terrestrial atmosphere and the solar nebula. Accretion of moonlets to form the Moon occurred in this region, possibly at a distance of $\sim 5 R_{\oplus}$ (Fig. 5).

It remains to consider the significance of the higher FeO content of the Moon compared with the Earth's mantle^{54,75}. Actually, the Mn (ref. 24), Co (ref. 53) and W (ref. 33) abundances in the Moon may also be slightly but significantly higher than in the Earth's mantle (Fig. 4). It has been suggested^{33,77} that the excess FeO was derived from planetesimals involved in the major impacts. This explanation might apply also to the excesses of Mn, Co and W. The planetesimals accreting at a late stage on the Earth may well have been oxidized, containing no free metal phase⁷⁸. The excess Fe, Mn, Co and W might be explained if the impact ejecta included $\sim 20\%$ of the projectile. Under the high temperatures pertaining, a limited degree of autoreduction is likely to have occurred, yielding $\sim 1\%$ of Ni-rich metal phase which ultimately segregated to form the lunar core.

1. Darwin, G. H. *Phil. Trans. R. Soc.* **170**, 447–530 (1879).
2. Jeffreys, H. *Mon. Not. R. astr. Soc.* **91**, 169–173 (1930).
3. Jeffreys, H. *Mon. Not. R. astr. Soc. Geophys. Suppl.* **4**, 62–71 (1937).
4. Urey, H. C. *The Planets* (Yale University Press, New Haven, 1952).
5. Urey, H. C. in *Physics and Chemistry of the Earth*, Vol. 2 (eds Ahrens, L., Press, F., Rankama, K. & Runcorn, S.) 46–76 (Pergamon, London, 1957).
6. Urey, H. C. *Yb. Phys. Soc. Lond.* **14**, 29 (1957).
7. Anders, E. A. *Rev. Astr. Astrophys.* **9**, 1–34 (1971).
8. Grossman, L. *Geochim. cosmochim. Acta* **36**, 597–619 (1972).
9. Ringwood, A. E. *Geochim. cosmochim. Acta* **15**, 257–283 (1959).
10. Ringwood, A. E. *Geochim. cosmochim. Acta* **30**, 41–104 (1966).
11. Ringwood, A. E. *Origin of the Earth and Moon* (Springer, New York, 1979).
12. Garz, T. *et al. Nature* **223**, 1254–1256 (1969).
13. Ringwood, A. E. *Geochim. cosmochim. Acta* **20**, 241–259 (1960).
14. Wise, D. U. *J. geophys. Res.* **68**, 1547–1554 (1963).
15. Cameron, A. G. W. *Icarus* **2**, 249–257 (1963).
16. O'Keefe, J. A. *J. geophys. Res.* **74**, 2758–2767 (1969).
17. Ringwood, A. E. *Earth planet. Sci. Lett.* **8**, 131–140 (1970).
18. Ringwood, A. E. *Phys. Earth planet. Inter.* **6**, 366–376 (1972).
19. Ringwood, A. E. *Composition and Petrology of the Earth's Mantle* (McGraw-Hill, New York, 1975).
20. Wänke, H. & Dreibus, G. in *Tidal Friction and the Earth's Rotation* Vol. 2 (eds Brosche, P. & Sundermann, J.) 332–344 (Springer, Berlin, 1982).
21. Warren, P. H. & Wasson, J. T. *Proc. 10th Lunar planet. Sci. Conf.* **2**, 2051–2083 (1979).
22. Taylor, S. & Jakes, P. *Proc. 5th Lunar Sci. Cong.* **2**, 1287–1305 (1974).
23. Ganapathy, R. & Anders, E. *Proc. 5th Lunar Sci. Conf.* **2**, 1181–1206 (1974).
24. Ringwood, A. E. & Kesson, S. E. *The Moon* **16**, 425–464 (1977).
25. Sun, S. *Geochim. cosmochim. Acta* **46**, 179–192 (1982).
26. Jagoutz, E. *et al. Proc. 10th Lunar planet. Sci. Conf.* **2**, 2031–2050 (1979).
27. Newsom, H. & Palme, H. *Earth planet. Sci. Lett.* **69**, 354–264 (1984).
28. Morgan, J. W., Wandless, G., Petrie, R. & Irving, A. *Tectonophysics* **75**, 47–67 (1981).
29. Chou, C., Shaw, D. & Crockett, J. *J. geophys. Res. Suppl.* **88**, A507–A518 (1983).
30. Brey, G. & Wänke, H. *Lunar planet. Sci.* **14**, 71–72 (1983).
31. Rammensee, W., Palme, H. & Wänke, H. *Lunar planet. Sci.* **14**, 628–629 (1983).
32. Dreibus, G. & Wänke, H. *Lunar planet. Sci.* **10**, 315–317 (1979).
33. Wänke, H. & Dreibus, G. in *Origin of the Moon* (Lunar and Planetary Institute, Houston, in the press).
34. Ringwood, A. E. *Proc. R. Soc.* **395**, 1–46 (1984).
35. Dreibus, G. *et al. Lunar planet. Sci.* **13**, 186–187 (1982).
36. Dreibus, G. & Wänke, H. *Proc. 27th int. Geol. Congr.* **11**, 1–20 (1984).
37. Laul, J. C. *et al. Geochim. cosmochim. Acta* (in the press).
38. Stolper, E. *Earth planet. Sci. Lett.* **42**, 239–242 (1979).
39. Wood, C. & Ashwal, L. *Proc. 12th Lunar planet. Sci. Conf.* **2**, 1359–1375 (1981).
40. Pepin, R. O. *Nature* **317**, 473–474 (1985).
41. Clayton, R. N. & Mayeda, T. *Proc. 6th Lunar Sci. Conf.* **2**, 1761–1769 (1975).
42. Clayton, R. N. & Mayeda, T. *Earth planet. Sci. Lett.* **62**, 1–6 (1983).
43. Dreibus, G. & Wänke, H. *Z. Naturforsch.* **35a**, 204–216 (1980).
44. Delano, J. W. & Ringwood, A. E. *The Moon and Planets* **18**, 345–425 (1978).
45. Newsom, H. *J. geophys. Res. Suppl.* **90**, C613–C617 (1985).
46. O'Keefe, J. A. *Naturwissenschaften* **59**, 45–52 (1972).
47. O'Keefe, J. A. & Urey, H. C. *Phil. Trans. R. Soc.* **285**, 569–575 (1977).
48. Dreibus, G., Spettel, B. & Wänke, H. *Proc. 7th Lunar Sci. Conf.* **3**, 3383–3396 (1976).
49. Rammensee, W. & Wänke, H. *Proc. 8th Lunar Sci. Conf.* **1**, 399–409 (1977).
50. Wänke, H., Dreibus, G. & Palme, H. *Proc. 9th Lunar planet. Sci. Conf.* **1**, 83–110 (1978).
51. Wänke, H., Dreibus, G. & Palme, H. *Proc. 10th Lunar planet. Sci. Conf.* **1**, 611–626 (1979).
52. Newsom, H. in *Origin of the Moon* (Lunar and Planetary Institute, Houston, in the press).
53. Ringwood, A. E. & Seifert, S. in *Origin of the Moon* (Lunar and Planetary Institute, Houston, in the press).
54. Delano, J. W. in *Origin of the Moon* (Lunar and Planetary Institute, Houston, in the press).
55. Wolf, R. & Anders, E. *Geochim. cosmochim. Acta* **44**, 2111–2124 (1980).
56. Blackshear, W. & Gapcynski, J. P. *J. geophys. Res.* **82**, 1699–1701 (1977).
57. Yoder, C. F. *Abstr. Conf. on the Origin of the Moon*, 6 (Kona, Hawaii, 1984).
58. Toksoz, M., Dainty, A., Solomon, S. & Anderson, K. *Rev. Geophys. Space Phys.* **12**, 539–567 (1974).
59. Newsom, H. & Drake, M. *Geochim. cosmochim. Acta* **47**, 93–100 (1983).
60. Nakamura, Y. *et al. Proc. 7th Lunar Sci. Conf.* **3**, 3113–3121 (1976).
61. Ruskol, Y. L. in *Soviet-American Conf. on the Cosmochemistry of the Moon and Planets* (eds Pomeroy, J. H. & Hubbard, N. J.) 812–822 (NASA, Washington DC, 1977).
62. Smith, J. V. *Un. Sci.* **5**, 718–720 (1974).
63. Mizuno, H. & Boss, A. *Icarus* **63**, 109–133 (1985).
64. Harris, A. & Kaula, W. *Icarus* **24**, 516–524 (1975).
65. Hayashi, C. in *Fundamental Problems in the Theory of Stellar Evolution* (eds Lamb, D. & Schramm, D.) 113–128 (International Astronomical Union, 1981).
66. Hayashi, C., Nakazawa, K. & Nakagawa, Y. in *Protostars and Planets* Vol. 2 (University of Arizona Press, 1986).
67. Sekiya, M., Hayashi, C. & Nakazawa, K. *Prog. theor. Phys.* **66**, 1301–1316 (1981).
68. Hartmann, W. K. & Davis, D. *Icarus* **24**, 504–515 (1975).
69. Cameron, A. G. L. & Ward, W. R. *Lun. Sci.* **7**, 120–122 (1976).
70. Cameron, A. G. L. *Icarus* **62**, 319–327 (1985).
71. Stevenson, D. J. *Abstr. Conf. on the Origin of the Moon*, 60 (Kona, Hawaii, 1984).
72. Benz, W., Slattery, W. & Cameron, A. G. W. *Lunar planet. Sci.* **17**, 40–41 (1986).
73. Rubey, W. W. *Geol. Soc. Am. Spec. Pap.* **62**, 631–650 (1955).
74. Walker, J. C. *Precamb. Res.* **17**, 147–171 (1982).
75. Ringwood, A. E. in *Origin of the Moon* (Lunar and Planetary Science Institute, Houston, in the press).
76. Boslough, M. & Ahrens, T. J. *Lunar planet. Sci.* **14**, 63–64 (1983).
77. Delano, J. D. & Stone, K. K. *Lunar planet. Sci.* **16**, 181–182 (1985).
78. Turekian, K. & Clark, S. P. *Earth planet. Sci. Lett.* **6**, 346–348 (1969).
79. Seifert, S. & Ringwood, A. E. *Lunar planet. Sci.* **17**, 789–790 (1986).
80. Newsom, H. & Drake, M. *Geochim. cosmochim. Acta* **46**, 2483–2489 (1982).
81. Haughton, D., Roeder, P. & Skinner, B. *Econ. Geol.* **69**, 451–467 (1975).
82. Rammensee, W. thesis, Mainz Univ. (1978).
83. Kimuar, K., Lewis, R. & Anders, E. *Geochim. cosmochim. Acta* **38**, 683–701 (1974).
84. Hertogen, J., Janssens, M. & Palme, H. *Geochim. cosmochim. Acta* **44**, 2125–2143 (1980).
85. Wolf, R., Woodrow, A. & Anders, E. *Proc. 10th Lunar planet. Sci. Conf.* **2**, 2107–2129 (1979).

Primordial density fluctuations and the structure of galactic haloes

P. J. Quinn^{*†‡}, J. K. Salmon^{†‡} & W. H. Zurek[‡]

N-body models used to study the formation of structure in $\Omega = 1$ universes reveal that the mass profiles of the collapsed structures—galactic haloes—are intimately related to the power spectrum of initial, gaussian, density perturbations. In particular, flat rotation curves observed in disk galaxies are obtained only when the exponent of the power law $P(k) \sim k^n$ is, on the megaparsec scale, less than -1 , but not as small as -3 . These results have important implications for various cosmological models.

THE structures observed in the Universe on the galactic and larger scales are believed to have been seeded by primordial density perturbations¹. These perturbations were presumably imprinted much earlier by a process such as inflation²⁻⁴. Later, they were modified in the course of the radiation-dominated era (redshift $z > 2.5 \times 10^4 \Omega h^2$). Both this initial imprinting, and the more recent modifications depend sensitively on the composition of matter in the Universe and on the physics at very high energies. Therefore, an understanding of the formation of structure from initial, seed perturbations is essential for cosmology as well as for particle physics.

We now explore, by means of computer modelling, the formation of collapsed and virialized structures in Einstein-deSitter ($\Omega = 1$) universes. We focus on a class of model universes characterized by gaussian, scale-free density perturbations. Their power spectra:

$$P(k) = Ak^n \quad (1)$$

are completely specified by the normalization constant A and the exponent n . k is the wavenumber. The white noise case corresponds to $n = 0$. Scale-free perturbations in an $\Omega = 1$ universe result in hierarchical clustering⁵⁻¹⁰, which is both reasonably easy to analyse and often a good approximation of more complicated and more realistic models. We shall also simulate the formation of haloes in CDM (cold dark matter) universes¹¹⁻¹³. Indeed, we shall normalize our scale-free power spectra so that on a 'galactic scale' they have similar power to the CDM case. The influence of Ω on the structure of galactic haloes will be treated elsewhere (W.H.Z., P.J.Q. and J.K.S., manuscript in preparation).

Observations of rotational velocities in disk galaxies indicate that the bulk of the mass of these objects is invisible 'dark matter'. Moreover, this dark matter is distributed so that the rotational velocities of stars, H II regions and neutral hydrogen¹⁴⁻¹⁸ are approximately independent of radius out to distances near to and even well beyond the optical radius of the galaxy. This implies a density (ρ) of dark material that, at large radii, falls off like;

$$\rho(r) \sim r^{-2} \quad (2)$$

or alternatively, the mass profile:

$$M(r) \sim r \quad (3)$$

We will demonstrate that mass profiles of this kind can be obtained when the initial density perturbations have at least as much energy on the large scale as they have on the small scale,

$n < -1$. This conclusion has been anticipated analytically by the 'secondary infall' models proposed by Gunn and Gott¹⁹. In particular, using the secondary infall paradigm, Hoffmann and Shaham²⁰ have conjectured that the average density profile should be related to the power-law exponent n by

$$\rho(r) \sim r^{-\gamma} \quad (4)$$

where

$$\gamma = \frac{3(3+n)}{(4+n)} \quad (5)$$

for $n > -1$, and should relax to $\rho \sim r^{-2}$ for $-1 > n > -3$. Our results are in approximate agreement with these conclusions for $n > -2$, even though the process of formation of collapsed objects does not seem to conform with the secondary infall picture.

Computer simulations

A gaussian field of initial density perturbations can be reproduced by writing:

$$\begin{aligned} \Delta\rho(\vec{r}) &= \rho(\vec{r}) - \langle\rho\rangle \\ &= \sum_{\vec{k}} |\delta_{\vec{k}}| e^{-i\phi_{\vec{k}}} e^{-i\vec{k}\cdot\vec{r}} \end{aligned} \quad (6)$$

Here $|\delta_{\vec{k}}|$ are the absolute values of the complex amplitudes of all the modes and $\phi_{\vec{k}}$ are their phases. One can realize power spectrum $P(\vec{k})$ by choosing:

$$\delta_{\vec{k}} = \sqrt{2P(\vec{k})} R_{\vec{k}}^{\delta} \quad (7a)$$

$$\phi_{\vec{k}} = 2\pi R_{\vec{k}}^{\phi} \quad (7b)$$

where $\{R_{\vec{k}}^{\delta}, R_{\vec{k}}^{\phi}\}$ are a pair of random numbers with distributions which are, respectively, gaussian in the interval $(0, \infty)$, and uniform in the interval $(0, 1]$.

We have found it extremely useful to implement different power spectra by using the same set of random number pairs. This allows us to model 'the same' fragment of the universe with different power laws and therefore to bring out systematic changes which would otherwise be overwhelmed by the random nature of the initial perturbations.

The density distribution can be faithfully represented in the form given by equation (6) only when, on the scales of interest, the amplitude $\sigma = \sqrt{\langle(\Delta\rho)^2\rangle}$ is much smaller than the average density;

$$\sigma \ll \langle\rho\rangle \quad (8)$$

This follows not just from the trivial requirement that ρ be positive; beyond the linear regime mode-mode interactions induce correlations between modes which in turn invalidates

* Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA.

† Theoretical Astrophysics, California Institute of Technology, Pasadena, California 91125, USA

‡ Theoretical Astrophysics, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

the assumption of gaussian density perturbations and makes the use of equation (6) impossible (see ref. 1).

We guarantee the validity of equation (6) by adjusting the normalization constant A in equation (1) so that mass fluctuations on the scale-of-interest (Λ) are small:

$$\left[\frac{\delta M}{M} \right]_{\Lambda} = 0.06 \quad (9)$$

We impose the same condition on $\delta M/M$ for the CDM models. It will be useful but will entail no loss of generality, to refer to this chosen scale Λ by analogy with the CDM case—for which A can be estimated by comparison with present-day galaxy correlation functions^{12,13,21}—as the ‘megaparsec scale’. In other words, we anticipate that ‘galactic haloes’ will form from Λ -sized fragments of the initial distribution after a time interval corresponding to the Hubble time. With this convention in mind we can now state that our models are initiated with the Hubble constant $H_0 = 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ at $z_{\text{int}} = 24$ and are evolved for $\sim 10,000$ – $14,000$ Myr beyond the conventional present which for this set of initial conditions occurs at an unrealistically short time of 6,700 Myr after the big bang. It is essential to keep in mind that for the power law models the above convention is nothing but a convention as the scale-free nature of $P(\vec{k})$ implies. Our results—density profiles of haloes—are independent of the value of H_0 . The Hubble constant influences only the total values of the masses of haloes, but not their shapes.

The initial conditions are imprinted on a system of 64^3 particles, each of them with a conventional mass of $m = 1.02 \times 10^9 M_{\odot}$, $\Omega = 1$. They represent a $(10 \text{ Mpc})^3$ section of the model universe. The initial conditions are imposed using the Zel’dovich ‘growing mode’ method^{1,20,22} which deforms the cubic lattice of particles in the desired manner. The evolution of the system to $z = 0$ is then simulated by a Fourier (cloud-in-cell) code with a 64^3 mesh.

The Fourier methods are inaccurate on scales smaller than a few mesh spacings. Therefore, we use it only to set up initial conditions for the N -body simulations. These are provided by the output of the Fourier code at $z = 5.25$ and/or $z = 10.1$. (We do not want to generate initial conditions this late in the history of the universe, as by then the inequality (8) is no longer valid and equation (6) cannot be trusted.) The use of the Fourier code allows us to employ sufficiently many particles at very early times to imprint initial conditions without the danger of introducing errors via shot noise or aliasing (ref. 21 and work in preparation). Moreover, using this technique we can obtain the first ‘blurred’ but, nevertheless, informative pictures of the model at $z = 0$.

Two kinds of N -body runs are used. Low-resolution simulations use particles with a mass $M_L = 27m$. A sphere with a diameter of 10 Mpc cut out from the Fourier cube provides the initial conditions. Each of the low resolution particles has the location and the velocity of the centre of mass of a $3 \times 3 \times 3$ fragment of the original Fourier lattice. However, particle masses are now close to $3 \times 10^{10} M_{\odot}$ which means that a typical galactic halo will contain no more than ~ 100 particles. Therefore, to investigate smaller scale structures we have employed a different mapping of the Fourier particles onto the N -body initial conditions.

High-resolution runs use a one-to-one mapping in a chosen sphere of 2 Mpc radius, and a very low-resolution mapping (64 Fourier particles to one N -body particle) towards the outskirts of the model. Masses in the transition region between 2 and 3 Mpc take on intermediate values. The resulting multi-resolution system is evolved using the N -body code for a time comparable to the corresponding low-resolution run (see Fig. 1).

Typically 5,000–7,000 particles are evolved and a run is completed within 2–4 hrs of Cray-1 CPU time. Usually more than one high-resolution sphere-of-interest is chosen for each low-resolution model. Therefore, the same fragment of the universe is modelled using three distinct resolutions. All of the N -body runs use a smoothing length of 10 kpc. Note that on the scales

on which all three resolutions are expected to be accurate, all of them yield comparable results. Low-resolution runs are then used to extract information about the large-scale structure, mergers, and so on. The results of these runs will be discussed elsewhere. Here we shall concentrate on the small-scale information contained in the high-resolution runs. Specifically, we shall consider the density profiles of haloes, that is objects with masses between $10^{11} M_{\odot}$ and $10^{12} M_{\odot}$.

Halo density profiles

Figure 2 contains a graphic summary of the key results discussed in this section. Two conclusions are immediately apparent: (1) the high-resolution portions of the particle plots in Fig. 2 contain related structures (the ‘same’ halo can be usually identified in the figures corresponding to the neighbouring values of n). (2) The realizations with more power on small scales (larger n) have larger numbers of more compact haloes. In particular, while the model with $n = -2.75$ exhibits only two reasonably well developed haloes, there are about 10 such structures of various sizes for $n = 1$. Moreover, models with a lot of power on the large scale exhibit more pronounced ‘caustics’ and ‘filaments’. The CDM model is the closest in appearance to the $n = -1$ and $n = -2$ cases. This is not unexpected: the effective exponent of the cold dark matter power spectrum is close to -1.5 on the megaparsec scale.

Most of the haloes in the high resolution region contain only high-resolution particles. However, some of them are contaminated by heavy particles that have entered from the low-resolution part of the model. The results given below are inferred from haloes that contain $\leq 10\%$ of these heavies by mass. (In fact, only 10% of the analysed halos had more than 1 or 2 heavies).

For the densities given by equation (4), the mass profile (the mass within radius r) is given by the relation:

$$M(r) \sim r^{3-\gamma} \quad (10)$$

The corresponding rotational velocity inferred for such $M(r)$ is;

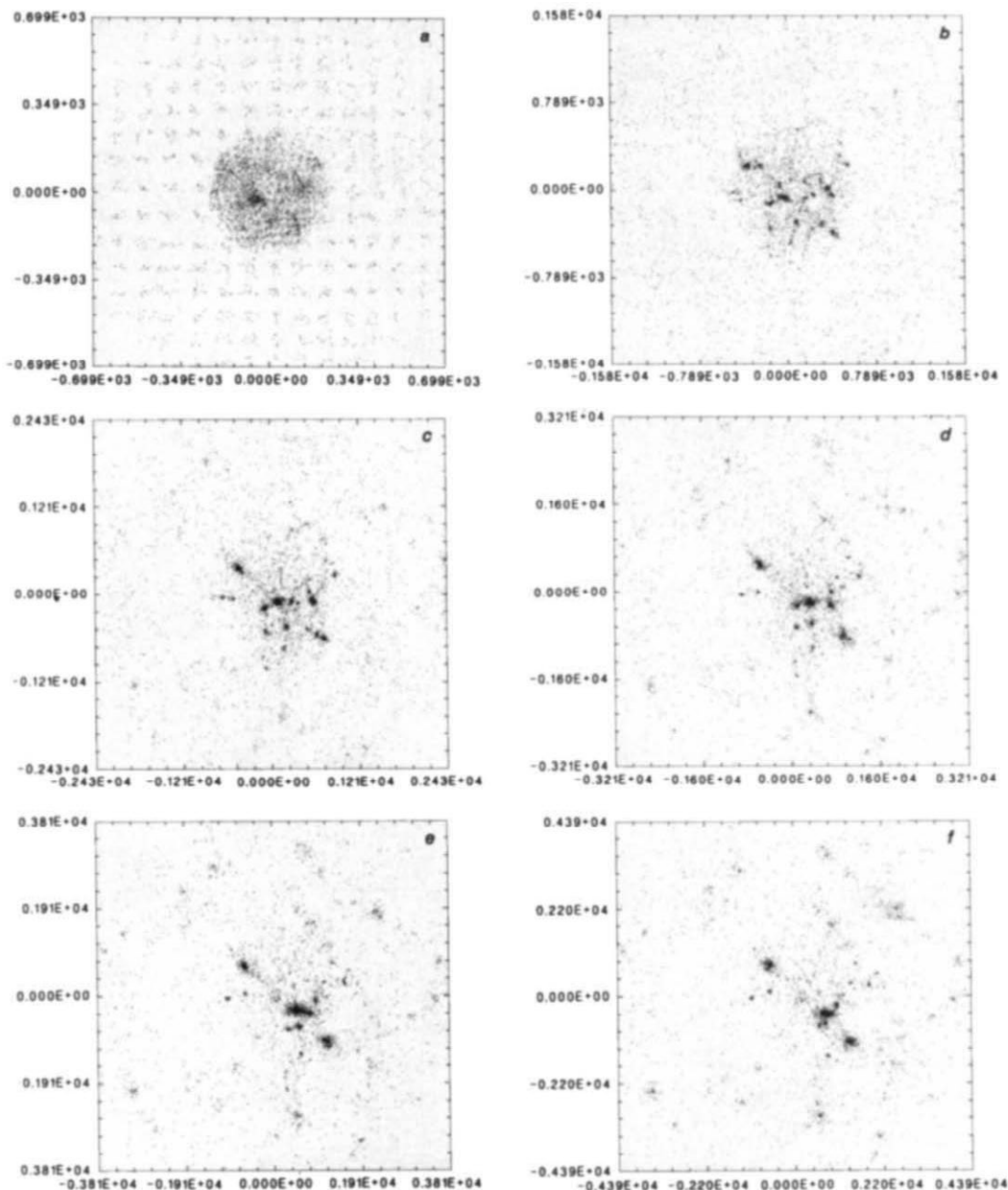
$$v(r) \sim r^{(2-\gamma)/2} \quad (11)$$

The circular rotational velocities of haloes obtained in the computer experiment are for a subset of the haloes shown in Fig. 2. We also indicate there the velocity profile predicted by equation (5).

Models with $n = 1$ and $n = 0$ produce centrally condensed haloes with the mass falling off faster than in the flat rotation case. The typical slope of the density profile is well approximated by the value of γ inferred from equation (5) although a systematic discrepancy can also be noted: densities tend to fall off faster than equation (5) would have it. In particular, rotation curves are not quite flat for $n = -1$ but clearly flatten for CDM as well as for the $n = -2$ case. In all cases when $n > -1$ rotation curves flatten at radii between 10 and 20 kpc. (The resolution of our N -body code is 10 kpc and does not allow us to comment on the structure on the innermost parts of the halo.) It is interesting that virtually all haloes formed in a simulation with a given power law have very similar rotation curves.

The situation is different when $n = -2.75$. Now the rotational velocities are still rising at $r = 50$ kpc and some of them do not flatten even in the 100 kpc range. This seems to imply that the structure of these haloes is inconsistent with the conclusions of Hoffman and Shaham, who conjecture that flat rotation curves will form even for $n = -3$. It is unlikely that such haloes with rising rotation velocities are a temporary, unstable phenomenon: We have checked the slope of the density profile at $z = -0.5$, at a time equal to 14,000 Myr. It is still rising and equation (5) still provides as good an estimate as a flat rotation profile. More relevant in understanding this discrepancy is probably the systematic increase of the size of the soft core with n decreasing to -3 . This trend is visible already for $n = -2$. Indeed, rotation profiles between 100 and 200 kpc, not shown in Fig. 2, are quite

Fig. 1 Six photographs taken in comoving coordinates at equal time intervals during the evolution of the high-resolution model ($P(k) \sim k^{-1}$). The initial conditions ($z = 5.25$) for the N -body run as obtained from the Fourier code (which was started with gaussian density perturbations at $z = 24$), are shown in *a*. Note that there are sizeable regions with large overdensities inside the high-resolution sphere, and a slightly perturbed lattice of massive particles on the outside. *b*–*e*, The state of the system at intermediate times. *f*, $z = 0$. Only the inner, high-resolution parts of such models were used (see Fig. 2). Units are kiloparsecs.



flat for $n = -2.75$. Therefore, the relevant part of Fig. 2 may be showing just the soft core parts of the haloes for this case. Moreover, the treatment given in ref. 20 breaks down for $n < -3$. Hence, it is not too surprising that its predictions for $n = -2.75$ are inaccurate. Indeed, one should be surprised that the arguments of Hoffman and Shaham give as good an estimate of the resulting density profiles as they do, as the assumptions on which they are based are only approximately valid.

The mass profile of the cold dark matter haloes is reasonably close to the flat rotation curve variety out to 100–150 kpc. Beyond that distance rotation profiles begin to fall off. The rotation parameters (λ) of the collapsed objects are typically in the 0.05–0.1 range ($\lambda = |E|^{1/2} J / GM^{5/2}$, where E is the total energy of the halo, J its angular momentum, M its mass, and G is the gravitational constant.) This is consistent with our earlier findings^{24,25} as well as with a number of other cold dark matter simulations²⁶.

We have concluded that the mass profiles of collapsed, virialized objects in the $\Omega = 1$ universe initiated with an effective power law spectrum on the galactic scale are determined over the range of interest by the value of n . Equations (4) and (5)

give a fair estimate of the corresponding density profiles. Although a power spectrum translates unambiguously into a rotation curve, one should keep in mind that the reverse argument is not so straightforward: The fact that a flat rotation curve is observed, often allows, within the errors, a range of halo density profiles. Moreover, baryonic material in the dissipative process of settling in the centre of the core will inevitably steepen the final rotation curve²⁷.

Comparisons with theory

Gunn and Gott have developed the secondary infall scenario based on the assumption that the formation of collapsed objects in an overdense region begins from the density peak which in due course accretes surrounding shells of material¹⁹. Each of these shells is thought to be much less massive than the already collapsed and virialized core. Assuming spherical symmetry, one can show that material from each consecutive shell will be eventually deposited at a radius:

$$r = \frac{r_m(\beta - 1)}{2} \quad (12)$$

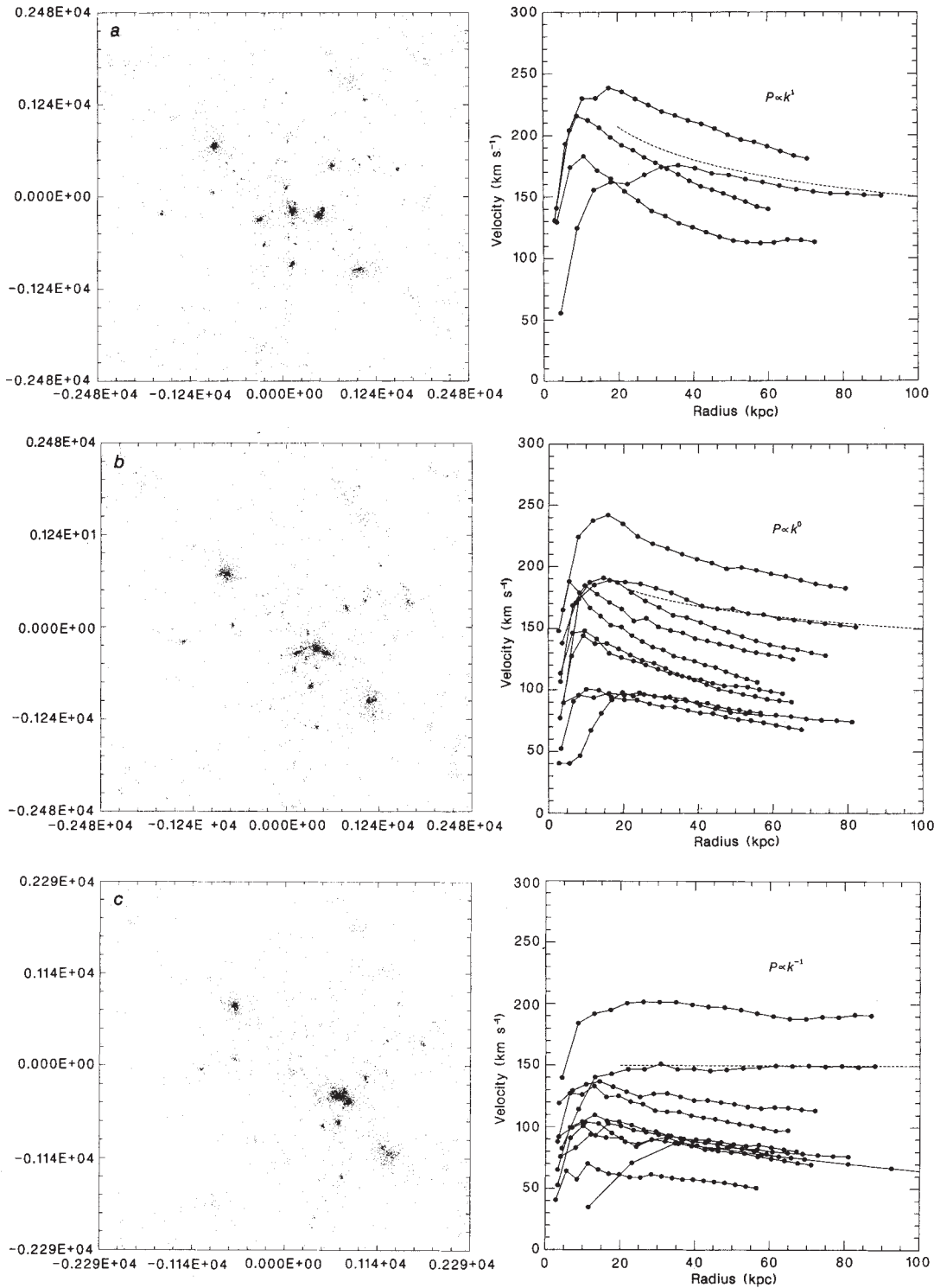


Fig. 2 The formation of haloes in an Einstein-deSitter ($\Omega = 1$) universe with different power spectra. Particle plots show only the inner, high-resolution regions. The exponent of the power law is indicated in the plots of the rotation velocities. The rotation profile predicted by $\rho(r) \sim r^{-3(3+n)/(4+n)}$ is indicated by dashed line. A scaled down average rotation profile of Sc spirals, ref. 15, is plotted along with the CDM haloes.

around the centre, where r_m is the radius of maximum expansion, and β characterizes the density of the already virialized core:

$$\rho(r) \sim r^{-\beta} \quad (13)$$

With a few additional assumptions, this can be used to calculate the density profile. The maximum radius of expansion for a spherically symmetric density perturbation with the initial relative overdensity δ_i is given by:

$$r_m = r_i \frac{1 + \delta_i(r_i)}{\delta_i(r_i) - (\Omega_i^{-1} - 1)} \quad (14)$$

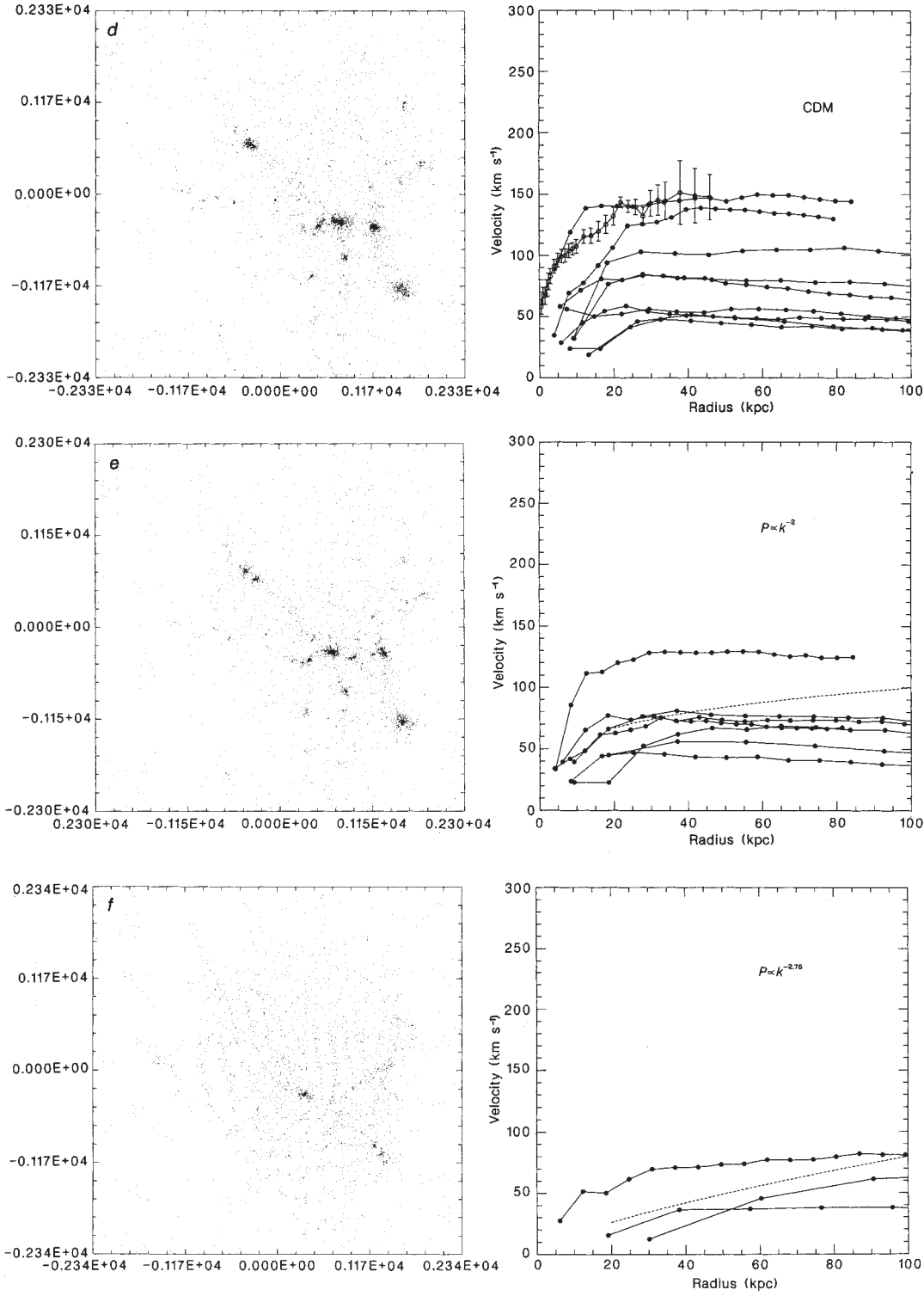
Here δ_i characterizes the initial overdensity inside the sphere of radius r_i :

$$M(r < r_i) = \left(\frac{4\pi}{3} \right) r_i^3 \langle \rho \rangle (1 + \delta_i(r_i)) \quad (15)$$

In the case considered here $\Omega_i = 1$. Therefore, the final density of the halo is:

$$\rho(r) \sim \delta_i(r_i(r))^3 \quad (16)$$

Hoffman and Shaham²⁰, following Peebles¹², argue that in a universe with a scale-free spectrum of initial density perturba-



tions, given by equation (1), the relative overdensity around a local maximum can be estimated by:

$$\delta_i \sim \left(\frac{r_i}{r_\lambda} \right)^{-(3+n)} \quad (17)$$

Here r_λ is the characteristic smoothing distance^{12,20,28}, related to the distance between the peaks. Consequently, the final density of the collapsed halo is given by:

$$\rho(r) \sim (r_i(r))^{-3(3+n)} \quad (18)$$

To express the right-hand side of this relation in terms of the distance from the halo centre we use equation (14) together with equation (17) to obtain:

$$r = \frac{r_m(\beta-1)}{2} \sim \frac{(\beta-1)}{2} \left(\frac{r_i}{r_\lambda} \right)^{4+n} \quad (19)$$

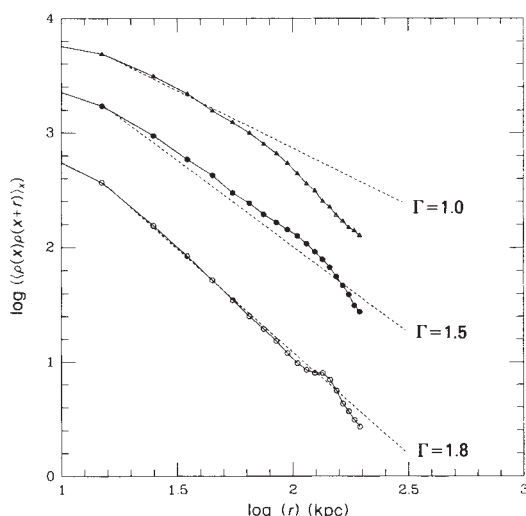


Fig. 3 Density autocorrelation functions, equation (21), for the particles in the high-resolution portions of some of the power law models. The prediction of the hierarchical clustering scenario—power law with the exponent given by equation (22)—is indicated by the dashed line. Results for: $\blacktriangle$, $n = -2$; $\bullet$, $n = -1$; $\circ$, $n = 0$ respectively. Note that each curve has been offset vertically by an arbitrary amount for clarity.

Therefore,

$$\rho(r) \sim \left(\frac{r}{r_\lambda}\right)^{-3(3+n)/(4+n)} \quad (20)$$

Our derivation of equation (20) paraphrases a more detailed discussion in ref. 20.

Hoffman and Shaham, following Fillmore and Goldreich²⁹, conjecture that the mass profiles with $M \sim r^{1+\epsilon}$, $\epsilon > 0$, are unstable. Therefore, they conclude that haloes with $\gamma < 2$ will relax into flat rotation curve haloes. This additional conjecture has not been so far confirmed by our computer models. The key difference between the situation analysed by Fillmore and Goldreich and our computer simulations may be due to the absence of angular momentum in the spherically symmetric, self-similar calculations reported in the ref. 29. Large angular momenta may stabilize haloes with rising rotation curves.

The analysis leading to equation (20) hinges on assumptions which are at best only approximately valid: (1) The initial overdensity profile, equation (17), is a reasonable estimate of the actual density profile only for a limited range of initial radii. When $r < r_\lambda$, near the peak, $\delta_i(r_i)$ is flatter than equation (17) would suggest. Moreover, for larger r , fluctuations of overdensity soon begin to exceed the systematic behaviour given by equation (17)^{20,28}. Therefore, the derivation is based on an assumption which is approximately accurate only for a rather limited range of radii. (2) A detailed discussion by Bardeen *et al.*²⁸ shows that the criterion implicitly used by Hoffman and Shaham to define density maxima is biased towards unusually broad peaks. (3) The density profile around correctly chosen density maxima falls off more steeply than equation (17) would have it. This difference may account for the discrepancy between the rotation curves predicted on the basis of equation (20) and the steeper rotation profiles obtained in our simulations. (4) The paradigm of the secondary infall onto a peak does not appear to be valid in the course of collapse leading to the formation of the haloes considered here. In particular, merging of smaller, already collapsed objects (see Fig. 1) appears to be a better description of the formation process. This is at odds with the picture of shells deposited spherically symmetrically onto a pre-existing core. In

spite of all these problems equations (4) and (5) give a reasonably accurate prediction of the density profiles of the collapsed objects. Discrepancies, which we have already pointed out in the previous section, are worth further study.

Scale-free spectra lead to a self-similar clustering hierarchy⁵⁻¹⁰. Peebles has pointed out that the density autocorrelation function in an evolved region of such a hierarchical universe should have a power law form⁶:

$$\langle \rho(x)\rho(x+r) \rangle_x = r^{-\Gamma} \quad (21)$$

with an 'hierarchical' exponent Γ given by:

$$\Gamma = \frac{3(3+n)}{(5+n)} \quad (22)$$

This Γ differs from the γ in the density formula, equation (5). We have verified (see Fig. 3) that this estimate for Γ is indeed correct in the high-resolution portions of our models. Note, perhaps not unexpectedly, that when the density autocorrelation is calculated only for the particles inside the haloes—that is, for x chosen from within the perimeter of the halo, but for an arbitrary r in the notation of equation (21)—it is considerably steeper, with the effective power law exponent close to the one given by equation (5).

Conclusions

Our numerical experiments provide clear evidence of a direct connection between the form of the power spectrum responsible for the gaussian density perturbations and the rotation curves of the resulting haloes. In particular, $P(k) \sim k^n$ with $n \sim -1 \rightarrow -2$ leads to flat rotation curves in present day galactic haloes. Power law spectra with $n > -1$ give haloes with rotation curves steeper than observed. This puts severe constraints on the shape of the power spectrum in the range corresponding to 1 Mpc. Furthermore, baryon condensation in the central parts of the halo can be expected to further steepen the rotation profile²⁷. This indicates that density perturbations with the preponderance of the power on small scales are untenable as models for galaxy formation. On the other hand, baryon condensation can probably flatten rising rotation curves inside broad soft cores, making $n = -1$ or $n = -2$ attractive. The cold dark matter power spectrum falls into the category of spectra producing haloes compatible with observations. Further studies to determine the environment dependence of other halo properties (such as the radii of their soft cores, angular momenta, and eccentricities) are underway.

We thank Stirling Colgate, Richard Epstein, Mike Fall, Yehuda Hoffman, Craig Hogan, Robert Hotchkiss, Nick Kaiser, Jeremy Ostriker and Martin Rees for stimulating discussion and help. Completion of this project would have been impossible without the large amounts of computer time provided by Los Alamos National Laboratory, Caltech and the Space Telescope Science Institute. We thank Kate Quinn for helping to prepare the manuscript.

Received 14 February; accepted 23 May 1986.

1. Peebles, P. J. E. *The Large Scale Structure of the Universe* (Princeton University Press, 1980).
2. Guth, A. *Phys. Rev. D* **23**, 347–356 (1981).
3. Linde, A. D. *Phys. Lett.* **108B**, 389–393 (1982).
4. Albrecht, A. & Steinhardt, P. J. *Phys. Rev. Lett.* **48**, 1220–1223 (1982).
5. Press, W. H. & Schechter, P. *Astrophys. J.* **187**, 425–438 (1974).
6. Peebles, P. J. E. *Astrophys. J.* **189**, L51–L53 (1974).
7. Gott, J. R. & Rees, M. J. *Astr. Astrophys.* **45**, 365–376 (1975).
8. Gott, J. R. & Turner, E. L. *Astrophys. J.* **216**, 357–371 (1977).
9. White, S. D. M. & Rees, M. J. *Mon. Not. R. astr. Soc.* **183**, 341–358 (1978).
10. Faber, S. M. in *Astrophysical Cosmology* (eds Bruck, H. A., Coyne, G. & Longair, M.) 191–217 (Pontificia Academia Scientiarum, 1982).
11. Peebles, P. J. E. *Astrophys. J.* **263**, L1–L5 (1982).
12. Peebles, P. J. E. *Astrophys. J.* **277**, 470–477 (1984).
13. Blumenthal, G. R., Faber, S. M., Primack, J. R. & Rees, M. J. *Nature* **311**, 517–525 (1984).
14. Rubin, V. C., Ford, W. K. & Thonnard, N. *Astrophys. J.* **238**, 471–487 (1980).
15. Rubin, V. C., Burstein, D., Ford, W. K. & Thonnard, N. *Astrophys. J.* **289**, 81–104 (1985).
16. Bosma, A. thesis, Univ. of Groningen (1978).
17. Carignan, C. & Freeman, K. C. *Astrophys. J.* **294**, 494–501 (1980).
18. van Albada, T. S., Bahcall, J. N., Begeman, K. & Sancisi, R. Preprint (Princeton Univ., 1985).

19. Gunn, J. E. & Gott, J. R. *Astrophys. J.* **176**, 1-19 (1972).
20. Hoffmann, Y. & Shaham, J. *Astrophys. J.* **297**, 16-22 (1985).
21. Efstathiou, G., Davis, M., Frenk, C. S. & White, S. D. M. *Astrophys. J. Suppl.* **57**, 241-260 (1985).
22. Zeldovich, Ya. B. *Astr. Astrophys.* **5**, 84-89 (1970).
23. Davis, M., Efstathiou, G., Frenk, C. S. & White, S. D. M. *Astrophys. J.* **292**, 371-394 (1985).
24. Quinn, P. J., Salmon, J. K. & Zurek, W. H. *IAU Symp.* **117** (in the press).
25. Zurek, W. H., Quinn, P. J. & Salmon, J. K. *IAU Symp.* **117** (in the press).
26. Frenk, C. S., White, S. D. M., Efstathiou, G. & Davis, M. *Nature* **317**, 595-597 (1985).
27. Blumenthal, G. R., Faber, S. M., Flores, R. & Primack, J. R. *Astrophys. J.* **301**, 27-34 (1986).
28. Bardeen, J. M., Bond, J. R., Kaiser, N. & Szalay, A. S. *Astrophys. J.* **304**, 15-61 (1986).
29. Fillmore, J. A. & Goldreich, P. *Astrophys. J.* **281**, 1-8 (1984).

The cytoplasmic carboxy terminus of M13 procoat is required for the membrane insertion of its central domain

Andreas Kuhn, William Wickner & Günther Kreil

Molecular Biology Institute and Department of Biological Chemistry, University of California, Los Angeles, California 90024, USA

The M13 coat protein spans the Escherichia coli plasma membrane with its amino-terminus facing the periplasm. It is made as a precursor—the procoat—with a typical leader peptide. Mutations which destroy the basic character of the carboxy-terminal domain of procoat, a domain which is oriented towards the cytoplasm, block membrane assembly, while insertion of three lysyl residues near the carboxy terminus partially restores assembly. Thus the information specifying membrane insertion of M13 procoat protein is found in its mature region as well as the leader and is not simply decoded in an amino to carboxy direction.

MANY secreted and membrane proteins are made with basic, amphiphilic leader (signal) peptides of 15 to 30 residues at their amino-terminus¹⁻³. Leader peptides are removed by an endoprotease, the leader peptidase⁴, after the pre-protein has crossed the membrane. Although leader peptides are essential for the processes of secretion and membrane assembly⁵, their precise roles have been the subject of speculation (reviewed in ref. 6) and are not yet established. Careful compilation and analysis^{2,3} have shown that certain features, such as the distribution of charged and apolar residues, are common among leader peptides. In contrast, there is no apparent sequence conservation among the mature regions of secreted proteins or extracytoplasmic domains of transmembrane proteins. This has led to the concept that the information specifying protein translocation into, or across, a membrane lies solely in leader peptides⁷.

Mutations in bacterial proteins which impair their export to the plasma membrane, periplasm or outer membrane can define domains which are relevant to this process. Almost all available mutations which affect the translocation of the λ receptor or the maltose-binding protein are in the leader region⁸. Furthermore, deletion of the carboxy-terminal regions of β -lactamase⁹ or maltose-binding protein¹⁰ has no effect on their export. These studies supported the concept that the leader peptide is sufficient to specify translocation.

However, other studies have implicated the mature region of pre-proteins in membrane transit processes. In bacteria, protein translocation is a late or post-translational event¹¹⁻¹³, suggesting that the carboxy-terminus may have a necessary role. Indeed, fusion proteins comprised of the leader peptide of a secreted protein and the carboxy-terminal region of a cytoplasmic protein cannot cross the plasma membrane^{14,15}. Furthermore, Silhavy and colleagues¹⁶ have shown that over half of the 'mature' region of the λ receptor must be present in a fusion protein to permit export. The relative roles of the leader and mature domains of a pre-protein during its export across the plasma membrane of *Escherichia coli* have thus remained undefined.

Membrane assembly of procoat

We have studied the biosynthesis of M13 coat protein, a small protein (relative molecular mass (M_r) 5,280) which spans the inner membrane of M13-infected cells¹⁷. It is made as a precursor, termed procoat (M_r 7,628), with a 23-residue leader peptide (Fig. 1). Our working model of procoat membrane assembly is shown in Fig. 1a. Procoat has two hydrophobic domains (drawn as rectangles in Fig. 1), one in the leader and one which spans the membrane in mature coat protein. Its membrane insertion is not coupled to either *secA* or *secY* function¹⁸ or to polypeptide synthesis¹⁹, but does require the transmembrane electrochemical potential²⁰. Transmembrane procoat is cleaved by leader peptidase to yield coat protein and the leader peptide, which is rapidly degraded. The biosynthesis of this protein has been studied extensively, both *in vivo* and in crude and reconstituted *in vitro* reactions^{21,22}. However, only a few mutants of this protein have been described.

In a recent study, we cloned the procoat gene and used *in vitro* mutagenesis with hydroxylamine to create mutations²³. More than 40 mutants were identified, of which 3 were in the leader sequence near the leader peptidase cleavage site²⁴. These three mutations affected the cleavage of procoat by leader peptidase, yet did not prevent procoat insertion across the membrane. In fact, despite careful screening, no mutations were identified which blocked membrane insertion. Examination of the 111 transitions which hydroxylamine could potentially cause in this gene of 219 base pairs (bp) revealed that only 68 would lead to amino-acid substitutions, and most of these would be quite conservative. Current mutagens only yield a very limited variety of mutants. Thus the failure to create export-defective mutations in the mature domain of secretory and membrane proteins with currently available mutagens cannot be taken as evidence that such mutations cannot occur.

In the present study, we have circumvented this problem by directly altering the carboxy-terminus of procoat with recom-

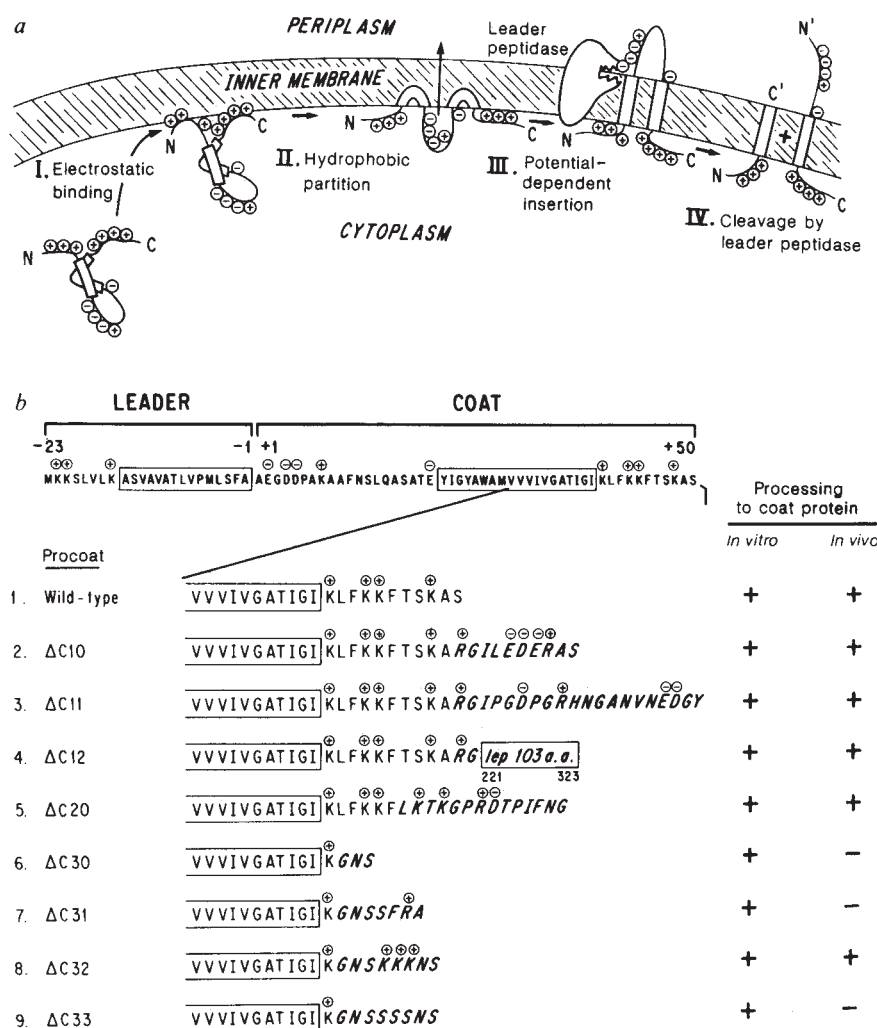


Fig. 1 *a*, A working model for the membrane insertion of M13 procoat. *b*, Construction of procoat mutants.

Methods. The procoat gene was cloned into the pING-1 vector²³ and excised by digestion with *Sall* and *EcoRI*. It was purified by electrophoresis in 0.7% agarose and then partially cleaved with *AluI*. This restriction endonuclease can cleave the procoat gene after the second base of the codon for the COOH-terminal serine and after the codon for lysine +40. The partially digested procoat gene was inserted into the pING plasmid which had previously been digested with *Sall* and *SmaI*. This yielded plasmids ΔC10 and ΔC30, respectively. As a second step, the plasmid ΔC10 was cut with *EcoRI* and ligated with the *EcoRI* fragment of leader peptidase. This contains the codons for the COOH-terminal region of leader peptidase, starting at isoleucine 221 and extending 52 bp beyond the stop codon. The resulting plasmids ΔC11 and ΔC12 contained this fragment in the opposite and same orientations as the leader peptidase gene, respectively. In ΔC12, the reading frame of the leader peptidase gene is retained, yielding a fusion protein. The same *EcoRI* fragment was also inserted into the plasmid ΔC30. This yielded only one phenotype, represented by plasmid ΔC31. Finally, the two complementary oligonucleotides d(AATTGGAAGAAGAAG) and d(AATTCTTCTCTCTCG) were synthesized and ligated into plasmid ΔC30 which had been digested with *EcoRI*. Depending on the orientation, this resulted in the insertion of three lysine (ΔC32) or three serine (ΔC33) residues. During screening for the clones ΔC11 and ΔC12, a new phenotype was observed which corresponded neither to the parent plasmid ΔC10 nor to ΔC12. This plasmid, termed ΔC20, was shown by sequencing to have a deletion which alters the amino-acid sequence of procoat after residue +45.

binant DNA techniques, including oligonucleotide-directed mutagenesis. We find that the basic character of the carboxy terminus of procoat is as essential for its membrane assembly as the leader peptide.

Carboxy-terminal mutations

The sequence of wild-type M13 procoat, and of mutations which we have created near the carboxy-terminus, are listed in Fig. 1b. The mutants are designated 'ΔC' to indicate the deletion of procoat near the carboxy terminus. The first digit after the C indicates which restriction site was used for the construction, while the second simply lists the various derivatives made at that site. All the ΔC constructions migrated at their expected M_r on SDS-polyacrylamide gel electrophoresis. Each was readily cleaved to its mature size in cell-free detergent extracts, as expected for the known substrate specificity of leader peptidase (ref. 25 and Figs 2, 4). We have shown previously that such cleavage is at the amino terminus, and is due to leader peptidase²⁴. In addition, the *in vitro* conversion of procoat to coat was shown to be inhibited by purified leader peptide (data not shown). Since these mutant procoat proteins are substrates for leader peptidase, their processing *in vivo* is an accurate indicator of whether they have assembled across the plasma membrane.

Procoat mutants ΔC10, ΔC11 and ΔC12 retained all but the carboxy-terminal serine of procoat, while adding 10, 20 or 105 polar amino-acid residues. Each of these mutant procoats rapidly chased to its respective mature form, though some of the precursor was lost to cytoplasmic degradation (Fig. 3, lanes

1, 2; 7, 8; 9, 10; odd-numbered lanes are pulse-labelled; even-numbered lanes are chased for 60 s). Since it has been suggested that the assembly properties of procoat are a function of its small size, we characterized the largest of these mutants, ΔC12, in detail. To create this mutant, a restriction fragment encoding the carboxy-terminal 103 polar amino-acid residues of leader peptidase was joined (in-frame) to the 3' end of the procoat gene. Its membrane assembly and orientation are indicated diagrammatically in Fig. 4a. *In vitro*, ΔC12 is synthesized as a precursor (Fig. 4b, lane 1) which can be cleaved to its mature M_r by leader peptidase (Fig. 4b, lane 2). *In vivo*, even a brief pulse-label reveals largely mature ΔC12 (Fig. 3, lane 9, and Fig. 4c) and only a small amount of precursor. This presumably simply reflects membrane assembly during the additional time needed to synthesize the 103 extra carboxy-terminal residues. Correct assembly was confirmed by protease mapping (Fig. 4c). Cells synthesizing ΔC12 were pulse-labelled with ³⁵S-methionine, then chilled and treated with Tris, sucrose and EDTA to permeabilize the outer membrane. These cells, or cells lysed with detergent, were digested with proteinase K for the indicated times. Aliquots were immunoprecipitated with anti-serum to coat protein or to leader peptidase and analysed by SDS-polyacrylamide gel electrophoresis (PAGE) and fluorography. Although a small fraction of the cells lysed during the experiment, the precursor form of ΔC12 was largely inaccessible to protease, though it was completely degraded when the cells were lysed with detergent. The mature form of ΔC12 was cleaved to a form of slightly lower M_r by proteinase K added to intact spheroplasts. This lower M_r form retained the carboxy-terminal

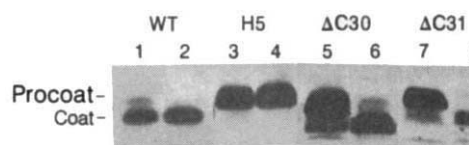


Fig. 2 Sensitivity of wild-type and mutant procoats to digestion by leader peptidase in detergent mixed micelles. *E. coli* HJM114-bearing plasmids with either the wild-type or various mutant procoat genes (as noted) were grown at 37 °C in minimal medium 9 (ref. 32) containing 0.5% fructose and a supplement of the mixed amino acids, less methionine. At an absorbance at 600 nm of 0.05, arabinose was added to 0.5% and growth continued for 2 h. Each culture (1 ml) was then labelled for 30 s with 50 μ Ci 35 S-methionine (1,000 Ci mmol $^{-1}$). Cultures were added to 0.25 ml of 30% sucrose, 0.6 mg ml $^{-1}$ lysozyme, 0.03 M EDTA, 3% (v/v) Triton X-100, and 100 μ g ml $^{-1}$ DNase I. Each sample was frozen in a dry ice/ethanol bath, then thawed and incubated for 30 min at 0 °C. Aliquots (0.2 ml) were either held on ice (lanes 1, 3, 5, 7) or incubated at 37 °C (lanes 2, 4, 6, 8) for 1 h, then analysed for procoat and coat by immunoprecipitation, SDS-PAGE and fluorography. Procoat H5 is a mutant in the leader peptidase target site of procoat²⁴.

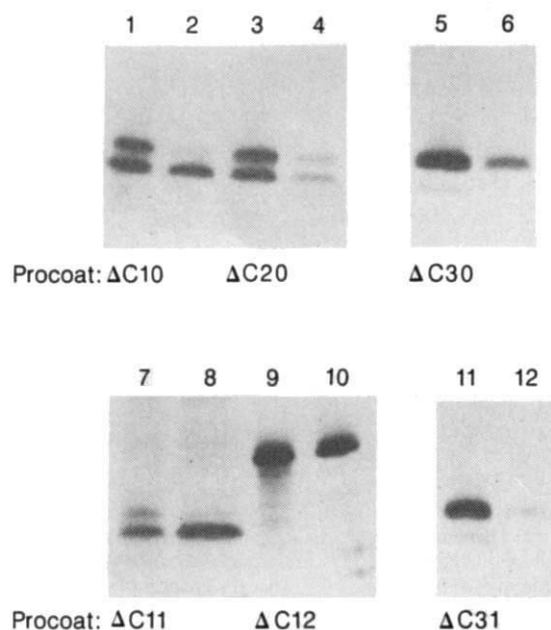


Fig. 3 Phenotypes of Δ C mutant procoats. Recombinant plasmids were transformed into wild-type *E. coli* HJM114. Cultures were grown overnight at 37 °C in M9 minimal medium containing ampicillin and glucose. Fresh overnight cultures were diluted 1:100 into M9 minimal medium containing 0.5% arabinose and grown at 37 °C to a density of 3×10^8 cells per ml. Aliquots were pulse-labelled for 30 s with 10 μ Ci of 35 S-methionine, then chased for either 5 s (odd-numbered lanes) or 60 s (even-numbered lanes) in the presence of unlabelled methionine. Samples were precipitated with trichloroacetic acid and analysed by immunoprecipitation, SDS-PAGE and fluorography.

leader peptidase immunological determinants, but was no longer immunoprecipitable by antiserum to coat protein (Fig. 4c). Δ C12 therefore has the expected orientation across the plasma membrane, with its amino terminus facing the periplasm and its carboxy terminus exposed to the cytoplasm. Mutants Δ C10 to Δ C12 demonstrate that simply elongating the carboxy terminus of procoat does not interfere with membrane assembly.

Procoat Δ C20 has lost the original carboxy-terminal 5 residues of procoat, including lysine +48, and in their place has sub-

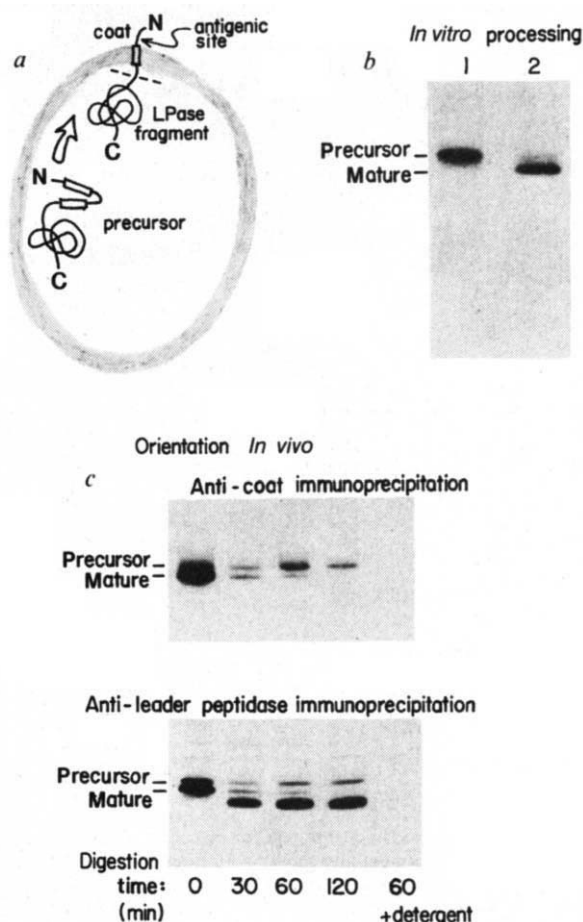


Fig. 4 Procoat Δ C12 is translocated across the membrane and processed. *a*, Schematic representation of the cellular localization of procoat Δ C12 and coat Δ C12. *b*, Plasmid-directed synthesis of procoat Δ C12 *in vitro* and its cleavage to a mature form by purified leader peptidase. *c*, Protease-accessibility of procoat and coat Δ C12.

Methods. Purified plasmid DNA encoding procoat Δ C12 was added to cellular extracts (S-30) of *E. coli* HJM114 in the presence of 0.4% arabinose, 0.5% Triton X-100 and purified ara C protein (200 μ g ml $^{-1}$; a gift from Dr G. Wilcox, Ingene Inc.). Incubation was at 37 °C for 1 h under the conditions described by Zalkin *et al.*³³ either in the absence (lane 1) or in the presence (lane 2) of 50 μ g ml $^{-1}$ purified leader peptidase⁴. For *c*, exponentially growing cells were pulse-labelled with 35 S-methionine for 1 min and osmotically shocked by mixing with an equal volume of ice-cold 60 mM Tris-HCl, pH 8.0, 40% sucrose, 20 mM EDTA. These cells were incubated with proteinase K (1 mg ml $^{-1}$) for the indicated times without (lanes 1–4) or with (lane 5) 2.5% Triton X-100. The samples were then immunoprecipitated with antibodies to coat (upper panel) or to leader peptidase (lower panel) and analysed by SDS-PAGE and fluorography.

stituted 14 polar, basic residues. It also assembles across the plasma membrane and is processed to coat protein (Fig. 3, lane 3), though this assembly reaction is inefficient. Δ C20 procoat was inaccessible to added proteinase K in spheroplasts, while the mature Δ C20 was degraded (Fig. 5a), confirming that processing reflected membrane assembly. Both the precursor and mature forms of Δ C20 were rapidly degraded *in vivo* (Fig. 3, lane 4). The precursor form is far more stable in a *htp*, *lon* strain (Fig. 6b), which lacks the cytoplasmic, ATP-dependent *lon* protease²⁶, indicating that procoat Δ C20 is normally degraded by the cytoplasmic *lon* protease in wild-type cells. The mature form of Δ C20 is still degraded, presumably by a different proteinase.

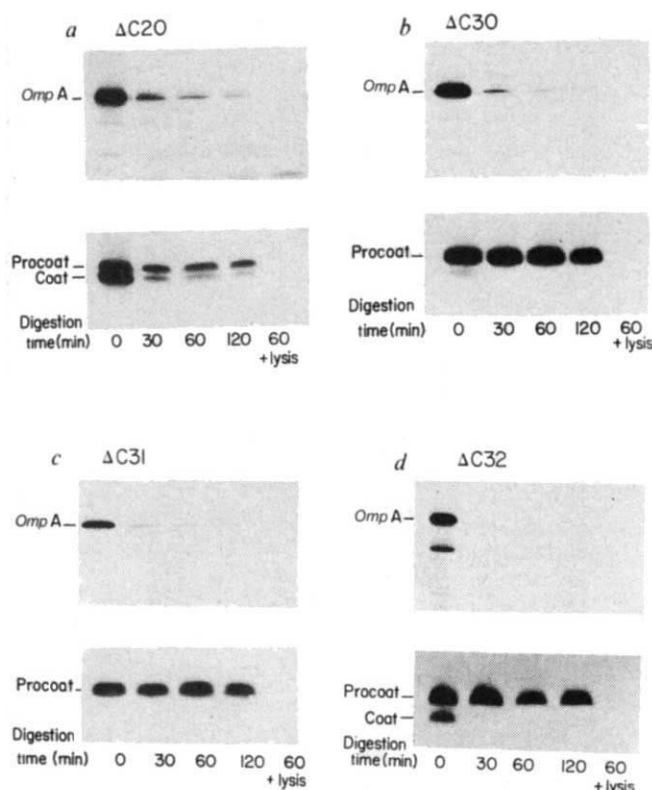


Fig. 5 Protease-accessibility to test for insertion of procoat into the membrane. Exponentially growing cultures of LC 137 which bore plasmids coding for procoat $\Delta C20$ (a), $\Delta C30$ (b), $\Delta C31$ (c), or $\Delta C32$ (d) were pulse-labelled with ^{35}S -methionine for 1 min, chased with an excess of L-methionine for 1 min, and osmotically shocked by thorough mixing with an equal volume of ice-cold 60 mM Tris-HCl pH 8.0, 40% sucrose, and 20 mM EDTA. Proteinase K (Boehringer) was added to a final concentration of 1 mg ml^{-1} and incubated at 0°C for the indicated times. For some samples, the cells were lysed by addition of 2.5% Triton X-100. Proteinase K was inactivated by addition of phenylmethanesulphonyl fluoride (Sigma) and 20% trichloroacetic acid. The acid precipitates were collected by centrifugation, twice resuspended in ice-cold acetone, sedimented and resuspended in $50 \mu\text{l}$ of 10 mM Tris-HCl (pH 8), 2% SDS and incubated for 5 min at 95°C . Each sample was then diluted with 1 ml of 2% Triton X-100, 0.1 mM EDTA, 0.15 M NaCl, 10 mM Tris-HCl pH 8.0 (buffer A) and mixed with $15 \mu\text{l}$ of a 10% suspension of formalin-fixed, SDS-washed *Staphylococcus aureus*¹¹ for 30 min. The *S. aureus* cells were removed by centrifugation. Antiserum to M13 coat protein and the outer membrane protein (*OmpA*) was added to the supernatant and incubated overnight. After addition of $15 \mu\text{l}$ of a 10% suspension of *S. aureus* cells, the samples were incubated for 1 h and centrifuged. The cells were sequentially suspended in buffer A and buffer A minus detergent and collected by centrifugation. They were then mixed with electrophoresis sample buffer, boiled and applied to 19% polyacrylamide gels in SDS containing 6 M urea. After electrophoresis, the gels were fixed, incubated in salicylate, dried and fluorographed¹¹.

Altering the carboxy terminus

In the mutant $\Delta C30$, the carboxy-terminal 10 residues of procoat were replaced by the sequence Gly-Asn-Ser. This yields a precursor which contains 66 rather than 73 residues and which lacks 3 of the 4 lysyl residues present in the cytoplasmic 'tail' of the wild-type protein. Surprisingly, this change at the carboxy terminus of procoat completely prevents its processing (Fig. 3, lanes 5, 6). Since procoat $\Delta C30$ is an excellent substrate for leader peptidase *in vitro*, the absence of *in vivo* processing indicates that it failed to insert across the plasma membrane.

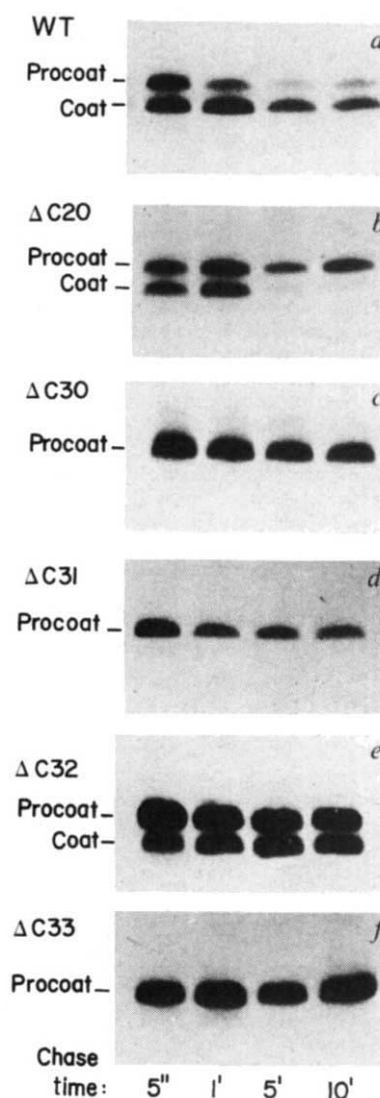


Fig. 6 *In vivo* processing of procoat mutants. Exponentially growing cultures of *E. coli* LC 137 (*htp*⁻, *lon*⁻) with plasmids coding for wild-type (a) or mutant (b-f) procoat were pulse-labelled with ^{35}S -methionine for 30 s and chased with an excess of L-methionine for the indicated times. Samples were analysed by immunoprecipitation, SDS-PAGE and fluorography.

The mutant precursor is slowly degraded in wild-type cells but not in the *htp*, *lon* mutant (Fig. 6c). This provides a second criterion that procoat $\Delta C30$ remains in the cytoplasm. Procoat $\Delta C30$ was also found to be inaccessible to protease in intact spheroplasts (Fig. 5b), confirming that it does not insert across the plasma membrane. These results cannot be explained by any combination of leader, stop-transfer and start-transfer sequences⁷; they are readily explained if procoat $\Delta C30$ fails to form a spontaneous insertion domain⁶.

A derivative of $\Delta C30$ was obtained by inserting the *EcoRI* fragment of the leader peptidase gene into the *EcoRI* site of the $\Delta C30$ gene. In either orientation, the fragment has only a short open reading frame, yielding procoat mutants of either 70 or 72 residues. Twelve such clones were investigated by pulse-chase experiments of the type described in Fig. 6, but the observed assembly phenotype was in each case the same as for $\Delta C30$. One such plasmid, $\Delta C31$, was sequenced and shown to encode a procoat mutant of 70 residues. The extra seven residues added after lysine +40 of the wild-type sequence are all hydrophilic and include one arginine. Although this precursor is nearly as

large as wild-type procoat and has two positive charges in its carboxy-terminal region, it still cannot translocate across the plasma membrane. It has the following properties: (1) it can be cleaved by leader peptidase in detergent extracts (Fig. 2, lanes 7, 8), but is not processed *in vivo* (Fig. 3, lane 11); (2) it is rapidly degraded in wild-type cells (Fig. 3, lane 12) but not in cells which lack the cytoplasmic *lon* protease (Fig. 6b); and (3) it is inaccessible to added protease in spheroplasts (Fig. 5c). Apparently, restoration of neither size nor overall polarity was sufficient to allow translocation of this mutant procoat.

We therefore devised a test of whether the basic charge at the carboxy terminus of procoat was essential for protein translocation. Two complementary oligonucleotides were synthesized which contained *EcoRI* linkers at their termini and, depending on the orientation, specified three codons for lysine or for serine. The two oligonucleotides were annealed and cloned into the *EcoRI* site of $\Delta C30$. The resulting procoat mutants showed different behaviour in pulse/chase experiments. Procoat $\Delta C32$, which has three lysines introduced near its carboxy terminus by the oligonucleotides, can insert into the plasma membrane (Figs 5d, 6e), though at a somewhat lower efficiency than wild-type procoat. In contrast, insertion of the oligonucleotide in the other direction, which created procoat $\Delta C33$ with three added seryl residues, did not restore membrane assembly (Fig. 6f). These results indicate that it is the basic charge in the short carboxy-terminal domain of procoat which is essential for its insertion across the membrane.

Conclusions

We have shown that the precise sequence of the carboxy terminus of procoat is not required for membrane assembly, yet its basic character is essential. The signal hypothesis, as originally proposed²⁷ and as modified⁷, postulates a vectorial decoding of linear information contained in 'signal' and 'stop-transfer' sequences of a membrane protein in the amino to carboxy direction. Our results demonstrate that the entire procoat protein has a role in the membrane assembly process. The carboxy terminus of procoat does not resemble a signal or stop-transfer sequence. It is highly polar, while apolar amino acids dominate signal or stop-transfer sequences. Since the carboxy terminus of procoat faces the cytoplasm, our results rule out a linear amino-to-carboxy decoding of this protein's assembly information. We suggest that the entire protein folds in a fashion competent for membrane insertion, as suggested by the membrane trigger hypothesis²⁸. Such a folded structure has recently been termed a spontaneous insertion domain⁶.

The carboxy terminus and leader peptide of procoat are not the only regions which are essential for its membrane assembly. We have also found that the introduction of an arginine at position +30 (the centre of the membrane-spanning domain) in place of the normal valine residue blocks membrane assembly (A.K., G.K. and W.W., manuscript in preparation). Thus the hydrophobic domain +20 to +40 is not only a 'membrane anchor', but is also required to initiate the membrane insertion process.

Our current working model of coat protein biogenesis is presented in Fig. 1a. Procoat may need its basic carboxy terminus in order to bind electrostatically to the membrane acidic lipid head groups. Both apolar domains, that in the leader and the apolar membrane anchor of the mature protein, may participate in partition into the apolar fatty acyl core of the membrane. Insertion of the central, polar domain across the membrane would then be driven by the membrane electrochemical potential, followed by cleavage to yield the mature coat protein. This speculative model may provide a framework for designing future studies.

Other proteins, with different assembly characteristics, may also require information encoded throughout their sequence for insertion across the bacterial plasma membrane. Procoat protein is representative of a class of proteins which do not require the *sec*-encoded functions for their export^{18,29}. We have also examined the biosynthesis of leader peptidase, an inner membrane protein which requires at least *secA* and *secY* function for its membrane assembly¹⁸. Leader peptidase spans the plasma membrane, with its amino terminus facing the cytoplasm and its carboxy terminus exposed to the periplasm³⁰. A polar, central domain which is carboxy-terminal to the membrane-spanning region is essential for the membrane assembly of leader peptidase (R. Dalbey and W.W., manuscript in preparation). Mutations have also been described in the mature regions of λ receptor¹⁶ and pre-maltose binding protein³¹ which affect their export. These results suggest that, in general, various domains of bacterial proteins in addition to the leader sequence may participate in the process of membrane assembly.

We thank Dr Satoshi Miura for performing the experiment shown in Fig. 6b, Dr Pierre Prentki for help with DNA sequencing and Marilyn Rice and Douglas Geisser for technical assistance. This work was supported by a grant from the NIH and a contract from Ingene, Inc. A.K. is a fellow of the Schweizerischen Stiftung fuer medizinisch-biologische Stipendien and the European Molecular Biology Organization. G.K. is a Visiting Professor of the UCLA Molecular Biology Institute.

Received 13 January; accepted 28 April 1986.

1. Milstein, C., Brownlee, G. G., Harrison, T. M. & Mathews, M. B. *Nature new Biol.* **239**, 117-120 (1972).
2. vonHeijne, G. *Eur. J. Biochem.* **133**, 17-21 (1983).
3. Perlman, D. & Halvorson, H. O. *J. molec. Biol.* **167**, 391-409 (1983).
4. Wolfe, P. B., Silver, P. & Wickner, W. *J. biol. Chem.* **257**, 7898-7902 (1982).
5. Carlson, M. & Botstein, D. *Cell* **28**, 145-154 (1982).
6. Wickner, W. T. & Lodish, H. F. *Science* **230**, 400-407 (1985).
7. Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1496-1500 (1980).
8. Michaelis, S. & Beckwith, J. A. *Rev. Microbiol.* **36**, 435-465 (1982).
9. Koshland, D., Sauer, R. T. & Botstein, D. *Cell* **30**, 903-914 (1982).
10. Ito, K. & Beckwith, J. *Cell* **25**, 143-150 (1981).
11. Ito, K., Date, T. & Wickner, W. *J. biol. Chem.* **255**, 2123-2130 (1980).
12. Koshland, D. & Botstein, D. *Cell* **20**, 749-760 (1980).
13. Randall, L. L. *Cell* **33**, 231-240 (1983).
14. Moreno, F. *et al. Nature* **286**, 356-359 (1980).
15. Kadosh, J. T. *et al. J. biol. Chem.* **259**, 2149-2154 (1984).
16. Hall, M. H., Schwartz, M. & Silhavy, T. J. *J. molec. Biol.* **156**, 93-112 (1982).
17. Wickner, W. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1159-1163 (1976).

18. Wolfe, P. B., Rice, M. & Wickner, W. *J. biol. Chem.* **260**, 1836-1841 (1985).
19. Date, T. & Wickner, W. T. *J. Virol.* **37**, 1087-1089 (1981).
20. Date, T., Goodman, J. M. & Wickner, W. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4669-4773 (1980).
21. Date, T., Zwizinski, C., Ludmerer, S. & Wickner, W. *Proc. natn. Acad. Sci. U.S.A.* **77**, 827-831 (1980).
22. Wickner, W. *Trends biochem. Sci.* **8**, 90-92 (1983).
23. Kuhn, A. & Wickner, W. *J. biol. Chem.* **260**, 15907-15913 (1985).
24. Kuhn, A. & Wickner, W. *J. biol. Chem.* **260**, 15914-15918 (1985).
25. Dierstein, R. & Wickner, W. *EMBO J.* (in the press).
26. Goff, S. & Goldberg, A. L. *Cell* **41**, 587-595 (1985).
27. Blobel, G. & Dobberstein, B. *J. Cell Biol.* **67**, 852-862 (1975).
28. Wickner, W. A. *Rev. Biochem.* **48**, 23-45 (1979).
29. Oliver, D. B. & Beckwith, J. *Cell* **25**, 765-772 (1981).
30. Wolfe, P. B., Wickner, W. & Goodman, J. *J. biol. Chem.* **258**, 12073-12080 (1983).
31. Bankaitis, V. A., Rasmussen, B. A. & Bassford, P. J. Jr *Cell* **37**, 243-252 (1984).
32. Miller, J. *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, New York, 1972).
33. Zalkin, H., Yanofsky, C. & Squires, C. L. *J. biol. Chem.* **249**, 465-475 (1974).

Detection of planetary systems and the search for evidence of life

Bernard F. Burke

Department of Physics, Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, USA

The Solar System is our only example of a planetary system, and the Earth is the only known instance of a planet harbouring life. The formation of planetary systems may occur commonly among stars like the Sun, but whether life arises frequently, rarely, or is a unique event is more problematical. Astrometric techniques capable of detecting major planets (similar to Jupiter, Saturn and Uranus) orbiting nearby stars are currently active¹⁻⁴, but detection of an Earth-like planet is beyond current techniques. Direct imaging of planetary systems (refs 5, 6; B. M. Oliver, unpublished data) is even more challenging, and current opinion seems to have dismissed the possibility of imaging a planet as small as the Earth. The problem is not an impossible one, however: by combining radioastronomy and optical methods, images of Earth-like planets orbiting nearby stars should be obtainable, and their atmospheres could be studied spectroscopically for evidence of life.

Consider the problem of optically detecting an Earth-like planet in orbit about the nearby G5 star τ Ceti, ~ 3 pc distant. Except for a numerical factor near unity, the signal-to-noise considerations are the same for both interferometric and filled-aperture systems. For a planet of diameter d , albedo η (assuming isotropic scattering), in a circular orbit of radius a at maximum elongation, the ratio of planetary to stellar photon flux (S_p/S_s) is $(\eta/16)(d/a)^2$. The flux from an Earth-like planet with $a = 1$ AU is indeed feeble: $(S_p/S_s) = 1.8 \times 10^{-10}$. More explicitly, the total photon flux of τ Ceti is 8.5×10^9 photons $\text{m}^{-2} \text{s}^{-1}$, and the scattered planetary flux would be 1.5 photons $\text{m}^{-2} \text{s}^{-1}$. Thus one must have a stable system that permits long integrations, and the flood of photons from the parent star must be reduced. The general system requirements, for a star at distance D whose flux would be S_0 if it were at distance D_0 , can be specified in terms of the effective collecting area A , the relative gain $G(\theta)$ for the star offset by an angle θ ($G=1$ at peak response) and signal-to-noise ratio r after integration time t : $A/G = r^2 (S_s/S_p)^2 (D/D_0)^2 / (tS_0)$. For $t = 10$ h, $r = 10\sigma$ (where σ is the standard deviation), and using the parameters for τ Ceti ($S_p/S_s = 1.8 \times 10^{-10}$, $D = D_0$ and $S_0 = 8.5 \times 10^9$ photons $\text{m}^{-2} \text{s}^{-1}$), detection of an Earth-like planet requires $(A/G) \approx 10^7 \text{ m}^2$. As any system would surely have a finite bandwidth, this must be a lower limit. Collecting area and side-lobe suppression are therefore key considerations.

Radio aperture-synthesis techniques have achieved remarkable success, starting from the earliest Cambridge work^{7,8}, and culminating in the Very Large Array (VLA)⁹. A study of the relationship between the optical and radio cases¹⁰ indicates that, with the signal-to-noise ratio treated in the shot-noise limit, as noted above, the radio concepts carry over to the optical domain. The ability to perform lengthy integrations (K. I. Kellerman *et al.*, personal communication) recently achieved a theoretical signal-to-noise ratio of 5σ after averaging for 76 h with the VLA) gives interferometry a definite advantage if coherent averaging is possible. It is natural, therefore, to consider extending the method to shorter wavelengths, and the optical case will be considered here, leaving infrared possibilities as a separate exercise. Optical interferometry on the ground is limited by 'seeing' fluctuations, and a space-based system would be free of such troubles. The possibilities for space-borne interferometry have received attention in the past year or so^{11,12}. The dispersed-fringe technique^{13,14} provides a solution of the 'white fringe' problem, and at the same time guarantees that every fringe-

visibility measurement retains information on the spectral distribution. For the present problem, the nearby star can provide a phase reference if the equipment is properly designed, thus providing coherent averaging. We therefore consider an interferometric system in space, having a total collecting area of 10 – 100 m^2 (the latter is admittedly ambitious). One can immediately see that the gain, G , must be between -50 and -60 dB. With a more realistic 10% bandwidth, the required gain must therefore be in the range -60 to -70 dB. This immediately rules out any interferometric system having uniformly illuminated apertures, because the side-lobe level would be much too high. The aperture illumination must therefore be tapered (apodized, in optical terminology). A simple example (not necessarily optimum) can be given by assuming a Hanning function taper, that is, one with amplitude proportional to $(1 + \cos(\pi r/r_0))/2$. The effective area is halved and the half-power primary beam width is increased, but the side-lobe level is decreased markedly.

If a planet like the Earth were orbiting τ Ceti, its maximum elongation would be 0.28 arcs. For this case, we assume an optical array of Hanning-tapered primary apertures, 1.5 m in diameter. With a desired signal-to-noise ratio of 10σ , a bandwidth that admits 10% of the photons, and an integration time of 10 h, it follows that a mean side-lobe level of -70 dB and a collecting area of 10 m^2 would be needed. This would be achieved by a 10-element array, using only photons of wavelength $\lambda > 5,000 \text{ \AA}$. This assumes, of course, that 'noise side-lobes' generated by optical imperfections are not important. Hence, one must be concerned with the optical quality of the system, and Ruze's theory¹⁵ for imperfect antennas applies. If the r.m.s. tolerance of the mirrors is $\lambda/60$ over distances ≥ 1 cm, a demanding but not impossible specification, the 'noise' side-lobes due to irregularities induced in the wavefront will be well below -70 dB, a conclusion consistent with the recent work of Mauron¹⁶. Detection of planets orbiting τ Ceti and other close stars, notably the dozen or so Sun-like stars within 6 pc, should therefore be possible.

We next examine spectrometric possibilities. The Earth's oxygen was almost certainly generated by life, probably by blue-green algae¹⁷, and a high oxygen abundance in the atmosphere of another planet would be a strong indication that life existed there. Owen¹⁸ proposed looking for the oxygen A-band absorption at $7,600 \text{ \AA}$, but he did not discuss how it could be done. He thought that Hartley ozone bands beyond $3,000 \text{ \AA}$ would be hard to distinguish from the Metopoulos-Beutler bands ($3,400$ – $2,600 \text{ \AA}$) of sulphur dioxide. With the spectroscopic resolution afforded by the dispersed-fringe correlator, however, this is not the case, provided the signal-to-noise ratio is sufficient. The ozone absorption, being at short wavelength, is easier to detect than the oxygen A-band because the side-lobe level is low. It appears that an integration time of ~ 100 h would give 10σ detection, and SO_2 , if present, could also be seen. Detection of the oxygen A-band is more delicate: the required resolution of $80 \text{ \AA}$ restricts the available photons, while, for fixed aperture size, the long wavelength dictates a raised side-lobe level. Nevertheless, it appears that an integration time of 100 h would give 10σ detection if the star could be kept within 1 marc s of the second null in the diffraction pattern.

Objections can be raised on several points, but there are ready replies. Scattered light, especially from zodiacal particles, might well be present. The signal level is so extended and weak that no problems arise for a zodiacal belt comparable to that of our Solar System. In reply to the objection that high photon efficiency has been assumed, note that present CCD (charge-couple device) detectors already have high photon efficiency. Read-out noise would be no problem because the background starlight effectively floods the field. Even if no further advances are made in the design of photon detectors, the required increase in integration time will be no more than a factor of two. Starlight scattered within the telescopes must be minimized, but careful optical design, as in a coronagraph, should avoid the problem.

The prospects are sufficiently encouraging to justify more intensive examination of interferometric systems in space. The present example has concentrated on the detection of Earth-like planets with familiar life forms, but in general one might expect that a planet teeming with life is likely to exhibit dramatic departures from chemical equilibrium in its atmosphere^{19,20}. Encouraging results from a successful instrument of the scale described in this note might well encourage more ambitious systems that might examine the ~ 50 F, G and K stars within 10 pc (J. C. Tarter, unpublished data). In the present discussion, the detection and study of Earth-like planets has been the principal aim; clearly, major planets could be detected easily and quickly by a similar interferometric system, and manifold opportunities would be opened in many other fields of astrophysical research.

I thank T. M. Donohue, F. D. Drake, A. K. Dupree, R. Giacconi, J. M. Moran (who provided apodization calculations), B. M. Oliver, M. Shao, R. D. Reasenberg and J. C. Tarter for helpful comments. B.F.B. was supported in part by a Smithsonian Regents Fellowships.

Received 10 September 1985; accepted 17 April 1986.

- Black, D. C. *Bull. Am. astr. Soc.* **16**, 767-769 (1984).
- Shao, M. *et al. Bull. Am. astr. Soc.* **16**, 750-757 (1984).
- York, D. G. *et al. Bull. Am. astr. Soc.* **16**, 775 (1984).
- Reasenberg, R. D. *Bull. Am. astr. Soc.* **16**, 758-766 (1984).
- Ken Knight, R. G. *Icarus* **36**, 422-423 (1978).
- Bracewell, R. N. & MacPhie, R. H. *Icarus* **38**, 136-147 (1979).
- Ryle, M. & Hewish, A. *Mon. Not. R. astr. Soc.* **120**, 220-230 (1960).
- Ryle, M. *Nature* **239**, 435-438 (1972).
- Napier, P. J., Thompson, A. R. & Ekers, R. D. *Proc. IEEE* **71**, 1295-1320 (1983).
- Burke, B. F. *Eur. Space Ag. spec. Publ.* **226**, 177-183 (1985).
- Reasenberg, R. D. (ed.) *Bull. Am. astr. Soc.* **16**, 747-837 (1984).
- Eur. Space Ag. spec. Publ.* **226** (1985).
- Labeyrie, A. in *Optical and Infrared Telescopes for the 1990s* (ed. Hewitt, A.) 786-788 (Kitt Peak National Observatory, Tucson, 1980).
- Stachnik, R. & Labeyrie, A. *Sky Telesc.* **67**, 205-209 (1984).
- Ruze, J. *Proc. IEEE* **54**, 633-640 (1966).
- Mauron, N. thesis, Univ. Marseilles (1984).
- Cloud, P. *Am. Scient.* **62**, 54-66 (1974).
- Owen, T. in *Strategies for Search for Life in the Universe* (ed. Papagiannis, M. D.) 177-185 (Reidel, Dordrecht, 1980).
- Margulis, L. & Lovelock, J. E. *Icarus* **21**, 471-480 (1974).
- Papagiannis, M. *IAU Trans.*, 323-329 (1982).

A space telescope for infrared spectroscopy of Earth-like planets

J. R. P. Angel, A. Y. S. Cheng & N. J. Woolf

Steward Observatory, University of Arizona, Tucson, Arizona 85721, USA

Owen¹ has reviewed the potential for detecting life on Earth-like planets of nearby stars, from atmospheric spectra. The presence of oxygen, as revealed, for example, by the 7,600-Å absorption band, would be of particular significance. But even the direct detection of a Jupiter-like planet around the nearest star is a formidable task, perhaps just possible with the Hubble Space Telescope^{2,3} or with a Michelson interferometer operating at 40-μm wavelength⁴. Earth-like planets, being fainter and closer in, are still more difficult. Here we show that a space telescope of 16 m diameter, apodized in a new way, could image and measure oxygen in the thermal infrared spectra of earthlike planets up to 4 pc away. Several interesting candidate stars lie within this distance.

Observation of our own Solar System from a distance of 4 pc would reveal the Earth at an angular distance of 0.25 arc s from the Sun, and nearly 10^{10} times fainter in visible light. Burke⁵ argues that a multi-element interferometer in space could be capable of optical detection and spectroscopy. As we discuss below, this would require mirrors of extraordinary precision. Figure 1 shows the advantage of observing in the infrared. The Earth's thermal flux (measured in photons) is about 20 times

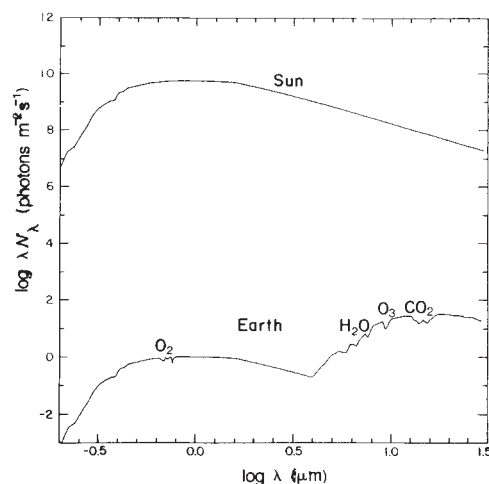


Fig. 1 Fluxes of the Sun and Earth as seen from 4 pc. Detection of spectral features in the photon noise limit depends on the photon flux N_λ in a given bandwidth, hence the quantity plotted is $\log(\lambda N_\lambda)$. The Sun's smoothed flux is from ref. 12. The Earth's optical flux assumes a constant albedo of 0.4 and isotropic emission with the O_2 band strengths from ref. 1. The infrared features are an average from ref. 6, normalized to be consistent with re-radiation of 60% of absorbed sunlight.

its optical reflected flux, while the solar flux at 10 μm is 35 times weaker than at 0.7 μm. As photon noise is a limiting factor, these are very significant advantages. Fortunately, the presence of oxygen results in a strong absorption feature (due to ozone) just below 10 μm, near the peak of the thermal spectrum. Figure 1 shows the main O_2 and O_3 features in the Earth's spectrum. An unambiguous detection of ozone needs a spectrum with a resolution of 1 μm and a signal-to-noise ratio of ~ 20 .

A telescope for optical or infrared imaging should form a diffraction pattern with very low side-lobes at the radius of the planet. The side-lobe intensities depend on the size and geometry of the telescope pupil, the wavelength λ , controlled variations in transmission or reflectivity (apodization) and phase errors across the pupil. In general, a mirror will have phase errors on different spatial scales l . At an angle θ from a star, the scattering will be determined predominantly by errors of wavelength $l \approx \lambda/\theta$. If the surface errors on this scale are random, with an r.m.s. amplitude equal to $\delta(l)$, then the maximum realizable gain G , the intensity at the diffraction pattern centre compared with that at θ , is given approximately by

$$G(\theta) \approx \frac{D^2 \theta^2}{8\pi \delta(l)^2} \quad (1)$$

where D is the telescope diameter. The exact value of G depends on the statistical character of the surface irregularities^{7,8}. The form of the residual scattered light will be a speckle pattern. Note that $G(\theta)$ depends only on the radiation wavelength through $\delta(l)$.

Consider first the problem of visible light imaging. The planet's image will be separated from that of the star by at least several times the radius of the first Airy minimum, depending on telescope aperture, and apodizing solutions have been developed in this domain capable of realizing extremely low-side-lobes. However, the tolerance on phase errors is very severe. Simply to overcome the photon noise of scattered starlight requires a gain of $\geq 10^7$ (ref. 5). Probably the most perfect mirror made to date is that of the 2.4-m-diameter Space Telescope, for which δ on the scale of 40 cm appropriate for visible light is ~ 5 nm (ref. 9). Thus, $G(0.25 \text{ arc s}) \approx 1.5 \times 10^4$, nearly 1,000 times smaller than required.

The only way to increase G is to decrease δ and/or increase D . For example, the goal of $G = 10^7$ could be achieved in an 8-m mirror if δ were 0.6 nm on a scale of 40 cm. The trouble is

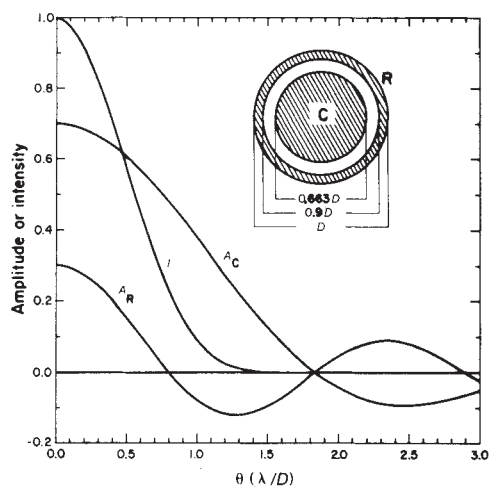


Fig. 2 Radial profiles of diffraction by a circular aperture apodized as shown, with a clear centre aperture C and outer annulus R. Curves A_C and A_R are the amplitudes from C and R, and I is the resulting intensity.

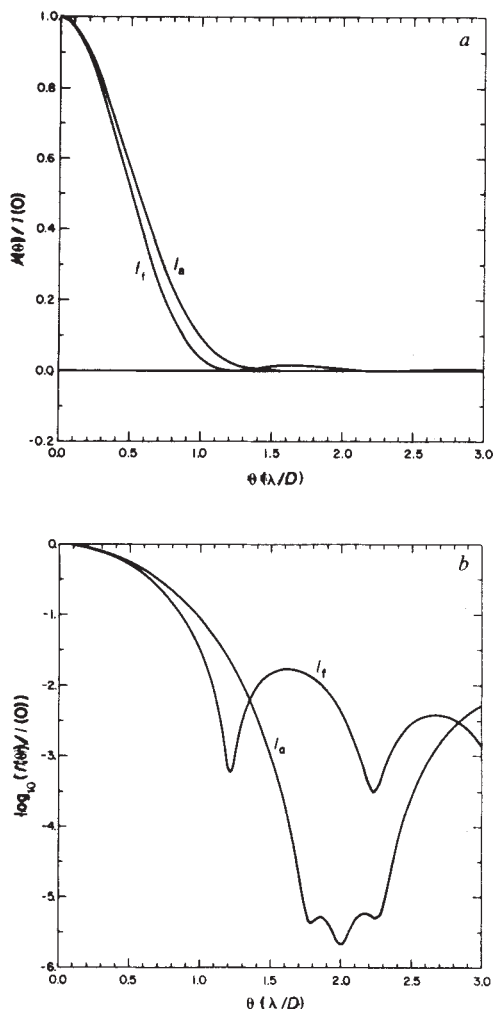


Fig. 3 Intensity of the diffraction patterns of the apodized aperture in Fig. 2 (I_a) and of a filled circular aperture (I_t) with the same outer diameter, plotted on different scales: *a*, linear; *b*, logarithmic. The intensities include broadening by bandwidth $\Delta\lambda/\lambda = 0.1$.

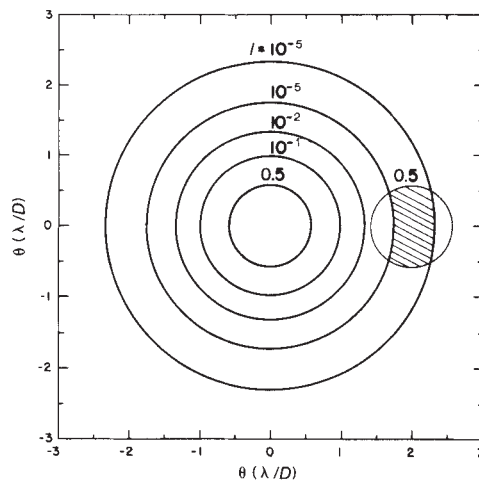


Fig. 4 Contour map of the intensity of the star's diffraction pattern, again with broadening by $\Delta\lambda/\lambda = 0.1$. The $I = 0.5$ contour of a planet's image is shown centred on the dark band, with the hatched area showing where the planet would be detected.

that even if the surface were finished to this accuracy, the remaining starlight is not a uniform background, but forms a speckle pattern. The speckles will look like a cluster of barely resolved stars of 21st magnitude, in which is hidden a 28th magnitude planet. Calibration of the 'star' field is almost impossible, as it is sensitive to path-length changes of only a few angstroms. These will probably arise as the telescope is rotated to move the planet's relative position or repointed to calibrate on another star. Burke⁵ argues that optical interferometry can control phase errors, but this will be a very difficult problem.

For these reasons we turn to the possibility of imaging the thermal emission, with greatly reduced requirements for gain and hence surface accuracy. The technical problems are still formidable, but if overcome will lead to a more secure observation. The telescope mirror must be large, at least 10 m in one dimension, to overcome the fundamental limits of diffraction at $10\ \mu\text{m}$ wavelength. The mirror must be cooled to $\leq 80\ \text{K}$, to reduce the telescope thermal background to a level below that of the zodiacal background and scattered stellar radiation. Such cooling may well be possible by largely passive thermal control in an appropriate orbit. For example, since the cryogen was exhausted, the IRAS (Infrared Astronomy Satellite) mirror has settled to a passive equilibrium temperature near 100 K (ref. 10).

The Airy diffraction pattern of the star formed by a simple round aperture of 10 m has a first minimum at 0.25 arc s radius, at $10\ \mu\text{m}$ wavelength. However, the dark ring is not a deep or broad enough null to allow an effective search for a planet. After consideration of various single-mirror and array configurations, we find the best search beam can be made by a circular aperture apodized in a new way. A very dark and broad first minimum would be produced if the wave amplitude at the first dark ring were made to have an inflection with zero gradient. This can be approximated accurately by masking a circular pupil with an obscuring annulus, as shown in Fig. 2. The second dark ring of the outer transmitting annulus is arranged to coincide with the first dark ring of the inner disk. By correct choice of the mask radii (0.663 and 0.9 of the full radius), the amplitude slopes can be made nearly equal and opposite, yielding a very dark ring at radius $\theta = 1.83\lambda/D$. For the planet at 0.25 arc s to appear centred in this ring at $10\ \mu\text{m}$ wavelength, the full mirror diameter must be 16 m.

Fine-tuning of the relative areas of the disk and ring components can yield a moderately high gain over a wide band, or very high gain over a narrower band. To determine which is better, consider the background of zodiacal emission. It would clearly be optimum to have the residual stellar flux weaker than

this background. The most recent estimate from IRAS of the zodiacal flux at the ecliptic pole is a grey body at 275 K with optical depth $\tau = 7.1 \times 10^{-8}$ (G. Rieke, personal communication). If the planetary system under observation is similar to the Solar System, and is viewed perpendicular to its orbital plane, its own zodiacal emission will be twice as bright, yielding a total background τ of 2.13×10^{-7} . The spatial resolution element $d\Omega$ of the apodized telescope is $(\lambda/D)^2$, and in this element, for $\lambda = 10 \mu\text{m}$ and $D = 16 \text{ m}$, we obtain a photon flux λN_λ of 850 photons $\text{m}^{-2} \text{s}^{-1}$. This is 5×10^{-6} of the $10\text{-}\mu\text{m}$ stellar flux of Fig. 1. Thus the gain should be significantly better than 10^5 , to avoid stellar emission contributing to the background shot noise.

Figure 3 shows a version of the apodization which gives a gain of 3×10^5 , averaged over a dark annular ring in the image plane $1.73 \lambda/D < \theta < 2.32 \lambda/D$. The amplitude passes through zero at three closely spaced radii. The profiles have been broadened by a bandwidth $\Delta\lambda/\lambda = 0.1$. The central maximum is only a little broader than for the unapodized case, while high throughput is maintained 63% that of the unapodized full aperture.

The signal-to-noise ratio of such a telescope can be computed with reference to Fig. 4, which shows the 50% contour of a planet's image seen against the star's diffraction pattern. Radiation from the planet would be measured with the shaded region of solid angle $d\Omega = 2.2 \times 10^{-13} \text{ sr}$, which contains 25% of its total flux. Assuming a combined telescope and detector efficiency of 50%, the signal from the Earth at 4 pc would be 30 electrons s^{-1} . Zodiacal emission would give a signal of 3,000 electrons s^{-1} , while the scattered light from the star in the same region would give a signal of 900 electrons s^{-1} . The shot noise fluctuations of the combined background would thus be ± 62 electrons s^{-1} , and the signal-to-noise ratio about unity in 4 s of integration. An integration of 30 min would be needed to achieve the desired signal-to-noise ratio of 20. We would use two-dimensional imaging detector with a tunable filter. The initial search would be made at one or a few wavelengths at the 0.25 arc s radius, and at larger radii with smaller or more conventional apodizing masks. Once a planet is located and an ephemeris calculated, its spectrum would be accumulated over a few hours by repeated scanning through the full wavelength interval.

Given an apodization solution that can, in theory, give very acceptable performance, we return to a discussion of mirror surface accuracy. For the 16-m mirror, phase errors on a scale of $l = 8 \text{ m}$ will be the most important, and from equation (1) for $G = 10^6$ these should be $< 4 \text{ nm}$ for the theoretical apodization gain of 3×10^5 not to be compromised. The mirror needs comparable accuracy to that of the Space Telescope on large scales, but could be less perfect on smaller scales. A 16-m mirror to these specifications could be assembled in space from lightweight rigid segments prefabricated on the ground. These should be made from glass that has zero expansion coefficient at the $\sim 80 \text{ K}$ operating temperature, such as high silica borosilicate with a conveniently low fusion temperature of $\sim 1,200^\circ \text{C}$ (ref. 11). The placement of the segments to form the large surface would have to be accurate to 4 nm. Such accuracy could be achieved by a servo loop to adjust the segment heights to correct errors in the star's diffraction pattern measured in ultraviolet light. This pattern is the same at different wavelengths, except that the effect of phase errors is magnified as the square of the wavelength ratio. Small phase errors in the $10\text{-}\mu\text{m}$ pattern will give error signals magnified by 1,000 at 3,000 Å and 10,000 at 1,000 Å, provided that spatial variation in ultraviolet and infrared reflectivity are controlled to the 0.1% level. The effect of 4 nm surface errors will be to introduce speckle structure in the dark ring at ~ 10 times the planet brightness. In the initial search this structure must be calibrated out by rotation of the telescope.

The search for earthlike planets and oxygen could conceivably be carried out with a less expensive, one-dimensional telescope configuration, but with inferior results. Consider, for example, a 4 m by 12 m rectangular mirror launched in the space shuttle

and apodized in a manner similar to that described, to give a response with two very dark fringes 0.25 arc s on either side of the central bright fringe. The planet could, in principle, be detected by rotation to compare the brightness of the two fringes, in the manner proposed by Bracewell and MacPhie⁴. However, it appears that the full dark ring formed by a round telescope will give a much better chance to reject spurious signals from phase errors, as well as more pixels. Furthermore, the inferior resolution solid angle of the smaller collecting area would result in a signal-to-noise ratio limited by zodiacal background which is four times worse. Each spectral point would thus take 16 times longer (8 h) and be less secure.

It is interesting that consideration of a single observational goal has led to a very specific solution: a round 16-m-diameter telescope, cooled to $< 80 \text{ K}$ and with a figure essentially diffraction-limited in visible light. Its size is similar to that proposed for the large deployable reflector (LDR) project, but it is cooled and of higher surface quality. It would bring great power to many of the most fundamental problems of astrophysics, in addition to the search for life. We believe it merits further study as part of NASA's long-range plans.

A corollary of this work is that life on Earth could have been detected during the past 10^9 years from thousands of planets beyond the wave of man-made radio emission, by an infrared telescope of 100–200 m diameter.

We acknowledge valuable discussions of this work with Bob Brown, Bernie Burke and Frank Low. Research in Innovative Optics for Space has been supported at Steward Observatory by NASA grant NAGW-121.

Received 11 November 1985; accepted 11 March 1986.

- Owen, T. *Strategies for the Search for Life in the Universe* (ed. Papagiannis, M. D.) 177–185 (Reidel, Dordrecht, 1980).
- Davies, D. W. *Icarus* **42**, 145–148 (1980).
- Breckinridge, J. B., Kuper, T. G. & Shack, R. V. *Opt. Enginng* **23**, 816–820 (1984).
- Bracewell, R. N. & MacPhie, R. H. *Icarus* **38**, 136–147 (1979).
- Burke, B. F. *Detection of Life in Other Planetary Systems*, *Nature* (in press).
- Rose, H. et al. *The Handbook of Albedo and Thermal Earthshine* (Environmental Research Institute of Michigan, Ann Arbor, Rep. No. 190201-1-T, 1973).
- Ruze, J. *Proc. IEEE* **54**, 633–640 (1966).
- Elson, J. M., Bennett, H. E. & Bennett, J. M. *Applied Optics and Optical Engineering Vol. 7* (eds Shannon, R. R. & Wyant, J. C.) pp. 191–244 (Academic Press, New York, 1979).
- Bahcall, J. N. & Spitzer, L. *Scient. Am.* **247**, 40–51 (1982).
- Mason, P. V. & Mount, K. *JPL Memo* (10 February 1984).
- Angel, J. R. P. *Proc. South-west Conference on Optics* (eds McDowell, R. S. & Gerada, J. B.) Proceedings of the SPIE, **542**, 32 (1986).
- Allen, C. W. *Astrophysical Quantities* (Clowes, London, 1973).

An explanation of the east–west asymmetry of Io's sodium cloud

Nicolas Thomas

Rock Cottage, Kinnerley, Oswestry SY10 8DF, UK

Since the discovery of a cloud of neutral sodium atoms accompanying the jovian satellite, Io, in its orbit¹, asymmetries of part of the cloud (region B)² have often been noted^{3–6}. Brightness variations with orbital longitude^{3,4} and asymmetries found by comparing images taken at orbital longitudes differing by 180° (ref. 5) have been explained as effects of solar radiation pressure⁷. Given that the vast majority of neutrals precede Io in its orbit, and that sputtering of Io is the principal source mechanism, Matson *et al.*⁸ postulated a hemispherical source centred 60° towards the sub-Jupiter point from the centre of the leading hemisphere to explain the general cloud shape. However, soft collisions with plasma ions may be important and call for a reassessment⁹. I show here that, given reasonable assumptions, the east–west brightness asymmetry can be explained by temporal changes in the number of emitting atoms in the cloud caused by atmospheric shielding of the surface.

Three types of SO₂ atmosphere have been proposed for Io. Matson and Nash¹⁰ used the phenomenon of sub-surface cold-trapping to produce a model in which sputtering of the entire surface is possible. Following measurements of sputtering yields¹¹, it has been shown^{12,13} that surface sputtering alone cannot balance observed loss rates from the plasma torus and hence cannot maintain the observed stability.

The second type of model¹²⁻¹⁴ is an atmosphere in which the sputtering ions cannot strike the surface but do sputter material from the atmosphere. In atmospheric sputtering, the bound gas serves as a target, and recoil collisions occur, leading to escape. Kumar himself¹⁴ recognizes the problems with this type of model. First, neutral sodium has to be present in large quantities in the atmosphere, where it is liable to rapid charge exchange. To reproduce the size of the cloud, a surface electron concentration 5 times higher than observed^{15,16} was found to be necessary in this model. Second, sputtering of the polar caps would not be possible, and rapid atmospheric mixing¹³ due to pressure gradients would lead to bright cap formation, contrary to observation¹⁷.

The third type of model^{18,19} is based on the equilibrium vapour pressure (EVP) of SO₂ controlling the atmospheric pressure. As Io's surface temperature varies from <87 K on the night side to >130 K at the sub-solar point, and the EVP of SO₂ increases by seven orders of magnitude over this range^{10,20}, the atmosphere is characterized by high pressures and atmospheric sputtering on the day side and low pressures and surface sputtering on the night side and at the poles. Thus at eastern elongation, 90° beyond superior conjunction, the co-rotating plasma strikes the night side and removes sodium and SO₂ from the surface, whereas at western elongation the plasma cannot strike the surface (except at the poles), and hence SO₂ in larger quantities is removed from the atmosphere. This idea implies an asymmetrical sodium source which should be reflected in changing numbers of atoms in the cloud and consequent brightness variations. Some authors, citing ref. 6, have suggested that the cloud is strikingly symmetrical between eastern and western elongations^{9,21}, and that therefore an anisotropic source is not appropriate. However, the work of Murcray and Goody⁶ is not conclusive. They state that "there may be a slight trend toward lower intensities in the exposures on the western side of Jupiter" (see also refs 3 and 4) and that they "are unable to judge from the data themselves whether or not we are seeing a real temporal change in the cloud". Thus an asymmetrical sodium source is not ruled out.

A simple model may be used to determine whether the idea is consistent with observations. The total ejection rate is taken as

$$\Gamma = kA \quad (1)$$

where k is the ejection rate per unit area. A is the cross-sectional area of Io exposed to the sputtering, as defined by

$$A = \frac{\pi r^2 - E}{2} (1 + \cos(\phi + \Phi)) + E \quad (2)$$

where

$$E = 4 \int_0^{\theta} r \sin(\pi/2 - \theta) (r^2 - y^2)^{1/2} dy \quad (3)$$

in which θ is the latitude above which no protection is afforded to the surface at any time, ϕ is the angular departure of Io from eastern elongation (Sun-centred coordinates) and r is the radius of Io. Φ (see Fig. 1) is an angle accounting for the process described by Cummings *et al.*²² which invokes neutral ionized near Io's surface as a source of sputtering ions. Interaction with the co-rotating electrical field would concentrate sputtering towards Io's inner face, thus explaining the relatively large number of sodium atoms inside Io's orbit. Inclusion of this term may be consistent with Io's longitudinal albedo asymmetry²³, showing the darkest hemisphere to be centred ~45° nearer the

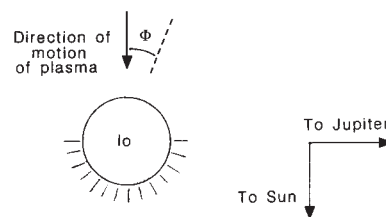


Fig. 1 At eastern elongation ($\phi = 0^\circ$), the higher-density day-side atmosphere (shaded) is sheltered from the incident plasma and affords no protection to the surface. The Φ term in the calculation accounts for the process described by Cummings *et al.*²².

sub-Jupiter point than the trailing hemisphere. Darker colour both on this hemisphere and at the poles is associated with radiation damage²⁴⁻²⁶ and lack of surface SO₂ due to surface sputtering²⁷⁻²⁹. In general accordance with these observations, Φ is taken to be $\pi/6$.

The equation for the total loss rate is taken as

$$\Xi = \frac{n(1 - e^{(-t/L)})}{t} \quad (4)$$

where L is the mean lifetime of sodium atoms in the cloud before ionization, n is the total number of atoms in the cloud and t is the time step. Thus the change in the total number of sodium atoms is

$$\frac{dn}{dt} = \Gamma - \Xi \quad (5)$$

Shemansky³⁰ calculated a value for L of 1 h; however, it has been suggested that this is only approached in the immediate vicinity of Io⁹. Arguments in favour of longer neutral sodium lifetimes are supported by high-resolution profiles of the sodium D₂-line showing high-velocity wings dependent on Io's orbital phase. Implied velocities are as high as 18 km s⁻¹ with respect to Io, blueshifted at eastern elongation and redshifted at western elongation³¹. A sputtered sodium atom would take just over 6 h to reach the leading edge of the cloud at this speed. Brown *et al.*⁹ consider that the average sodium lifetime in region B is between 5 and 10 h. They also point out that the velocity distribution and cloud structure can be explained by elastic collisions. The wings on the D₂-line profiles are strongest when Io is on Jupiter's magnetic equator³¹. As the peak ion density in the torus is on or near the magnetic equator³², the higher density should lead to more collisions and hence more sodium atoms at high velocity.

Figure 2, calculated by step-by-step integration of equation (5) (for $t \ll L$) with a shooting method, shows a brightness reduction consistent with 20-25% brightness asymmetries^{3,4}. Furthermore, the latitude above which surface sputtering can occur on the day side is consistent with that predicted by Pearl *et al.*'s¹⁸ EVP atmospheric model.

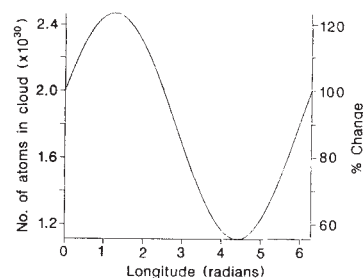


Fig. 2 Brightness variation as a function of orbital phase for Io's sodium cloud, assuming an anisotropic source caused by atmospheric shielding of the surface. Longitude is measured from eastern elongation. Day-side surface exposed to co-rotating plasma at latitudes greater than 75°; average sodium atom lifetime, 6.0 h.

Figure 3a shows the brightness reduction as a function of lifetime, maintaining the exposure latitude at 75° , and Fig. 3b shows the result of keeping the lifetime constant while varying the exposure latitude. Increasing the lifetime requires an increase in the exposure latitude to achieve the same brightness change. Thus sodium lifetimes and exposure latitudes are tightly constrained but are consistent with the EVP atmospheric models, surface brightness and SO_2 distributions and sodium velocity distributions. For example, Φ values $> \pi/4$ are inconsistent with observation. For $\Phi = \pi/6$, the lifetime is constrained to lie between 3 h ($\theta = 60^\circ$) and 8 h ($\theta = 90^\circ$). The lower limit is inconsistent with the velocity distribution. Near-direct sputtering of

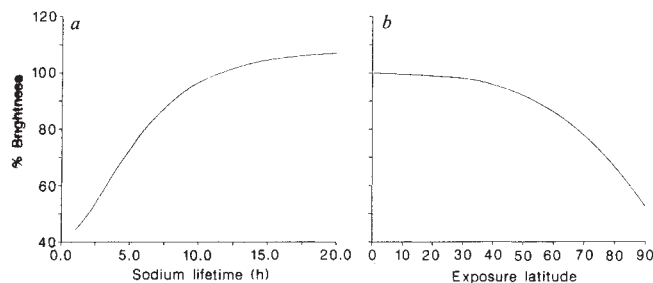


Fig. 3 Brightness changes at western elongation as a function of *a*, sodium atom lifetime (exposure latitude = 75°) and *b*, exposure latitude (mean sodium lifetime = 5 h). Brightness at eastern elongation = 100%.

the trailing hemisphere ($\Phi = 10^\circ$) gives a looser bracket for the lifetime of 10 ± 4 h.

Further temporal variations may result from volcanic activity, in which the driving volatile may protect the surface near the vent, leading to a short-timescale reduction in brightness of the cloud.

To estimate whether surface sputtering can produce sufficient quantities of sodium atoms to populate the cloud and, ultimately, the torus, knowledge of the sputtering yields is required; however, a rough value can be obtained by using the yields estimated for SO_2 (ref. 11). Taking the O^+ ion impact yield of 50 as a lower limit for the lower-mass Na atom and estimating that $\sim 3\%$ of the sputtered material would escape from Io, then 1.5 Na atoms per impact would escape from a pure sodium surface. Using $2 \times 10^9 \text{ m}^{-3}$ for the ion concentration in the plasma³², the number of sodium atoms produced would be $1.7 \times 10^{14} \text{ m}^{-2} \text{ s}^{-1}$. Multiplying by the cross-sectional area of Io gives a total of $\sim 1.8 \times 10^{27}$ Na atoms s^{-1} produced from one hemisphere if the surface were 100% sodium. Kumar¹⁴ states that the required rate to populate the cloud is 8.3×10^{25} atoms s^{-1} . This allows us to put a limit of $\sim 4.7\%$ on the percentage surface coverage of sodium. However, Kumar's required population rate is relatively low, a value of $\sim 10^{27}$ Na atoms s^{-1} having been quoted elsewhere⁹. Unless the sputter yield is substantially higher (by more than a factor of 4), surface sputtering would be capable of supplying this quantity only if a high proportion of the surface were composed of sodium and its compounds, although the higher value makes the alternative case for atmospheric sputtering of sodium even more difficult.

I thank Michael Woolfson and Beryl Leatham-Thomas for their assistance with this work.

Received 17 February; accepted 29 April 1986.

1. Brown, R. A. in *Exploration of the Planetary System* (eds Woszczyk, A. & Iwaniszewska, C.) 527-531 (Reidel, Hingham, 1974).
2. Brown, R. A., Goody, R. M., Murcray, F. J. & Chaffee, F. H. *Astrophys. J.* **200**, L49-L53 (1975).
3. Bergstralh, J. T., Matson, D. L. & Johnson, T. V. *Astrophys. J.* **195**, L131-L135 (1975).
4. Bergstralh, J. T., Young, J. W., Matson, D. L. & Johnson, T. V. *Astrophys. J.* **211**, L51-L55 (1977).
5. Goldberg, B. A., Garneau, G. W. & LaVoie, S. K. *Science* **226**, 512-516 (1984).
6. Murcray, F. J. & Goody, R. M. *Astrophys. J.* **226**, 327-335 (1978).
7. Smyth, W. H. *Astrophys. J.* **234**, 1148-1153 (1979).
8. Matson, D. L., Goldberg, G. A., Johnson, T. V. & Carlson, R. W. *Science* **199**, 531-533 (1978).

9. Brown, R. A., Pilcher, C. B. & Strobel, D. F. in *Physics of the Jovian Magnetosphere* (ed. Dessler, A. J.) 197-225 (Cambridge University Press, 1983).
10. Matson, D. L. & Nash, D. B. *J. geophys. Res.* **88**, 4771-4783 (1983).
11. Johnson, R. E., Garret, J. W., Boring, J. W., Barton, L. A. & Brown, W. L. *J. geophys. Res.* **89**, B711-B715 (1984).
12. Kumar, S. *J. geophys. Res.* **89**, 7399-7406 (1984).
13. Cheng, A. F. *J. geophys. Res.* **89**, 3939-3944 (1984).
14. Kumar, S. *Icarus* **61**, 101-123 (1985).
15. Kliore, A. J. *et al. Icarus* **24**, 407-410 (1975).
16. Kliore, A. J. *IAU Colloq.* **57** (1980).
17. Minton, R. B. *Commun. lunar planet. Lab.* **10**, 35-39 (1973).
18. Pearl, J. *et al. Nature* **280**, 755-758 (1979).
19. Fanale, F. P., Banerdt, W. B., Elson, L. S., Johnson, T. V. & Zurek, R. W. in *The Satellites of Jupiter* (ed. Morrison, D.) 756-781 (University of Arizona Press, Tucson, 1982).
20. Honig, R. E. & Hook, H. O. *RCA Rev.* **21**, 360-368 (1960).
21. Kumar, S. & Huntten, D. in *The Satellites of Jupiter* (ed. Morrison, D.) 782-806 (University of Arizona Press, Tucson, 1982).
22. Cummings, W. D., Dessler, A. J. & Hill, T. W. *J. geophys. Res.* **85**, 2108-2114 (1980).
23. Nash, D. B. & Johnson, T. V. *Icarus* **38**, 69-74 (1979).
24. Moore, M. H. *Icarus* **59**, 114-128 (1984).
25. Radford, H. E. & Rice, F. O. *J. chem. Phys.* **33**, 774-776 (1960).
26. Young, A. T. *Icarus* **58**, 197-226 (1984).
27. Soderblom, L. A. *et al. Geophys. Res. Lett.* **7**, 963-966 (1980).
28. Howell, R. R., Cruikshank, D. P. & Fanale, F. P. *Icarus* **57**, 83-92 (1984).
29. Nelson, R. M. *et al. Science* **210**, 784-786 (1980).
30. Shemansky, D. E. *Astrophys. J.* **236**, 1043-1054 (1980).
31. Trafton, L. & Macy, W. *Astrophys. J.* **215**, 971-976 (1977).
32. Bagenal, F. & Sullivan, J. D. *J. geophys. Res.* **86**, 8447-8466 (1981).

A class of narrow-band-gap semiconducting polymers

Samson A. Jenekhe

Honeywell Inc., Physical Sciences Center, Bloomington, Minnesota 55420, USA

Scientific interest in electrically conducting polymers and conjugated polymers in general has been widespread among workers in polymer science, chemistry, condensed matter physics, materials science and related fields since the discovery of doped conductive polyacetylene^{1,2}. Many doped conducting organic polymers with conductivity spanning the range from insulator to near-metallic ($\sim 10^{-15}$ – $10^3 \text{ ohm}^{-1} \text{ cm}^{-1}$) are now known¹⁻¹³. Of prime importance and fundamental interest in the continuing experimental and theoretical search for new conducting, and perhaps superconducting, polymers is the achievement of small or vanishing values for the semiconductor band gap (E_g), which governs the intrinsic electronic, optical and magnetic properties of materials. Existence of a finite E_g in conjugated polymers is thought to originate principally from bond-length alternation, which is related to the Peierls instability theorem for one-dimensional metals¹⁴⁻¹⁷. Here I describe a novel class of conjugated polymers, containing alternating aromatic and quinonoid segments, whose members exhibit intrinsic band gaps as low as 0.75 eV, the smallest known value of E_g for an organic polymer.

Among conjugated polyenes, polyenynes, and related polymers, polyacetylene (PA), $(-\text{CH}=\text{CH}-)_n$, has the smallest band gap, with a value of 1.5 eV (ref. 3). Among the aromatic and hetero-aromatic conjugated polymers, polythiophene (PT) (Fig. 1a with X = S) has the smallest band gap, with a value of ~ 2.1 – 2.2 eV (refs 8, 13). In the search for narrower-band-gap polymers, the strategy of substitution on existing main-chain polymers or formation of co-polymers has not generally been successful. An important exception is the annelated 3,4-benzo derivative of polythiophene, polyisothianaphthene (PITN), which has an E_g value of only 1.13 eV. Wudl *et al.*¹³ have attributed the difference in E_g values between PT and PITN to contribution of quinonoid resonance structure (Fig. 1b) in the case of PITN and a lack of such contribution in PT¹³. I have argued elsewhere that a large part of the ~ 1 eV difference in E_g values of PT and PITN is attributable to co-planarity of polymer repeating units in PITN and its lack in PT¹⁸. However, introduction of quinonoid character into the main chain of an aromatic conjugated polymer can be expected to reduce the band gap.

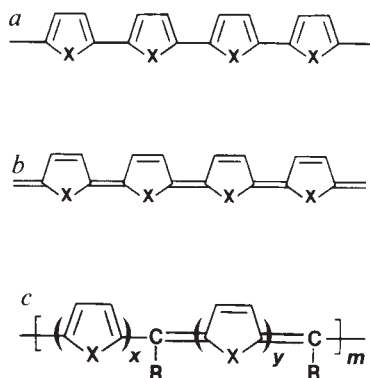


Fig. 1 *a*, Aromatic ground-state structure of hetero-aromatic polymers (for example, X = S, polythiophene; X = N-H, polypyrrole). *b*, Quinonoid ground-state structure of hetero-aromatic polymers of *a*. *c*, Ground-state structure of novel tunable, narrow-band-gap, conjugated polymers containing alternating aromatic and quinonoid segments, some of which exhibit the smallest E_g values known for organic polymers.

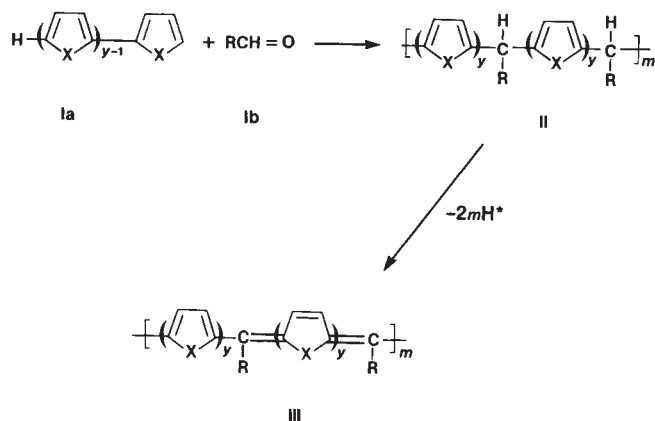


Fig. 2 Two-step synthesis of polymers with alternating aromatic and quinonoid segments, with quinonoid character $Q = 1/2$ ($x = y$ in Fig. 1c).

In fact, we have already observed a similar effect in polycarbazoles¹⁰. Also, Bredas¹⁹ has recently shown theoretically, using valence effective hamiltonian (VEH) calculations, that as quinonoid structure is introduced into polythiophene geometry (Fig. 1*a, b*) E_g decreases linearly with increasing quinonoid character¹⁹.

Figure 1*c* shows a class of hetero-aromatic conjugated polymers designed to incorporate aromatic and quinonoid segments in the main chain, which we hoped would exhibit small E_g values. This class of polymers displays several interesting general features:

(1) The quinonoid character is given in terms of integer molecular parameters (x, y) which could be subject to synthetic manipulation. Neglecting possible end-group effects, the fraction of the chain with quinonoid character is $Q = y/(x+y)$. Thus, $Q = 0, 1/2, 2/3, 3/4, 4/5, \dots$ when $y/x = 0, 1, 2, 3, 4, \dots$ and $Q = 1/3, 1/4, 1/5, 1/6, \dots$ when $y/x = 1/2, 1/3, 1/4, 1/5, \dots$. The limiting cases of $Q = 0$ ($y = 0$ or $x \rightarrow \infty$) and $Q = 1$ ($x = 0$ or $y \rightarrow \infty$) correspond respectively to polymers with chain structures *a* and *b* (Fig. 1). The hypothetical polymer *b*, relative to *a* and polymers of intermediate compositions, has a lower but nonetheless finite E_g ¹⁹. The parameter Q could in principle be related to bond-length alternation Δr ($=d_1 - d_2$) by some function $f(\Delta r)$, where d_1 and d_2 are the C—C and C=C bond lengths respectively. The value of Q when $\Delta r = 0$ and of Δr when E_g is minimum would be of interest in understanding the origin of E_g in conjugated polymers.

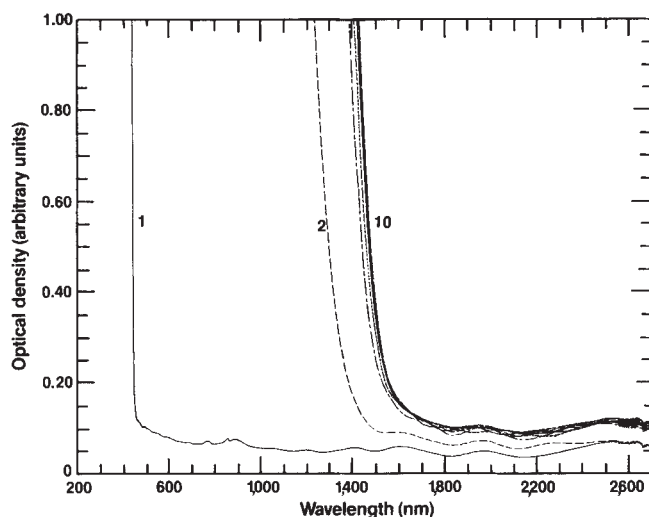


Fig. 3 Optical absorption spectra of thin films of polymer precursor PTTB (1), and conjugated derivatives (2–10), which are the polymers in Fig. 1*c*.

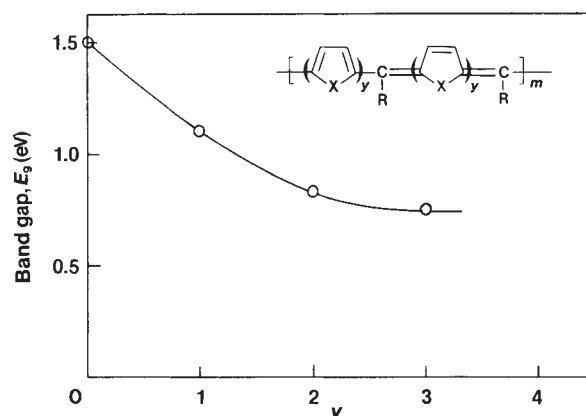


Fig. 4 The band gap $E_g(y)$ of the polymers in Fig. 1*c* with X = S, R = Ph and $x = y$. The E_g value at $y = 0$ is that of *trans*-PA.

(2) The ground-state structure of this class of polymers is two-fold degenerate in the sense of *trans*-PA, when $Q = 1/2$ or $x = y = 1, 2, 3, 4, \dots$. In fact, when $x = y = 0$ ($Q = 0$) and R = H we obtain polyacetylene; however, the uniqueness of polyacetylene is preserved by its $Q = 0$ value. Soliton excitations have been postulated for *trans*-PA on the basis of its degenerate ground-state structure²⁰. Thus, soliton excitations may be anticipated in the degenerate or symmetric ($Q = 1/2$) polymers of Fig. 1*c*. However, it is unclear to what extent a finite Q will influence such excitations relative to *trans*-PA.

(3) If the hetero-atom (X) is neglected, the polymer backbone is essentially an alternating co-polymer of *trans*-cisoid and *cis*-transoid forms of polyacetylene.

(4) Doping in aromatic (for example, poly-*p*-phenylene) and hetero-aromatic (Fig. 1*a*) polymers leads to polarons and bipolarons as the charged species, presumably by increasing the quinonoid character (Q) of the chain^{21,22}. As the present class of polymers already has $Q > 0$, it is uncertain how much more, if any, quinonoid character can result from doping.

(5) As macromolecules, these polymers have an unusual 'expandable' constitutional repeating unit (CRU), consisting of alternating aromatic and quinonoid segments. Conceivably, a material with a very high molecular weight (say 10^7) could result from a small chain length (say $m = 1$), simply by increasing x and y . If such materials could be synthesized this suggests that the electronic and optical properties would be more sensitive to x and y than to the polymer chain length m . These

features and questions warrant a detailed theoretical investigation.

Polymers of the type shown in Fig. 1c, with $x = y = 1, 2$ or 3 ($Q = 1/2$) have been experimentally realized for various X and R by a two-step synthetic technique, shown in Fig. 2. First, non-conjugated precursor polymers containing alternating sp^3 carbon atom ($-CRH-$) and hetero-aromatic conjugated units D were synthesized by condensation polymerization of the appropriate monomer, dimer or trimer (Ia) with aldehydes (Ib), where D is 2,5-thiophenediyl, 5,5'- α -bithiophenediyl or 5,5'- α -terthiophenediyl. Second, the precursors (II) were converted to the conjugated polymers with alternating aromatic and quinonoid segments (III) by oxidative elimination of the bridge hydrogens. This two-step synthesis could also be used to achieve the asymmetric polymers of Fig. 1c; that is, those with $Q \neq 1/2$. Note that the basic linear structure of the conjugated polymers (III) is the same as that of the precursors (II); the structure of the precursors (II) was readily established by infrared spectra. The precursor polymers have been characterized by infrared and electronic spectra, elemental analysis, studies of molecular weight, and thermal analysis. Evidence of the elimination of the bridge hydrogens is provided principally by infrared and electronic spectra; insolubility of polymers III compared to polymers II provides additional proof of elimination. Details of the syntheses of the conjugated polymers and their precursors are described elsewhere (refs 18, 23 and S.A.J. in preparation). The smallest band gaps were obtained with $X = S$. Also, E_g values were less sensitive to the side-group R .

Figure 3 shows the optical absorption spectra for thin films (~ 0.1 – $0.2 \mu m$) of a polymer with $R = \text{phenyl}$, $X = S$ and $x = y = 3$. Curve 1 is the optical absorption spectrum of the polymer precursor, poly-(5,5', α -terthiophenediyl benzylidene) (PTTB). Curve 2 is that of a conjugated derivative (Fig. 1c) which has an E_g of 0.83 eV (1,500 nm) and corresponds to the product of elimination after 7.1 min of reaction. On exhaustive elimination (curves 2–10) the band gap narrows to a constant value of 0.75 eV (1,650 nm). Similar results of band-gap narrowing to a constant value during elimination have been obtained for $x = y = 1$ and 2: $y = 1$, $E_g = 1.1$ eV; $y = 2$, $E_g = 0.83$ eV.

Figure 4 shows the band gap as a function of y for $X = S$ and $x = y$. The point shown at $y = 0$ is that for *trans*-PA; this is because the general polymer repeating unit shown in Fig. 3 contains $8y + 2$ carbon atoms in the main chain, which naturally reduces to polyacetylene in the limit $y = 0$. The asymptotic value of $E_g(y)$ is ~ 0.7 eV, which is already closely approached at $y = 3$.

As expected, the proposed polymers do exhibit narrow band gaps. Although the quinonoid character Q qualitatively explains why E_g is smaller in this class of polymers, relative to PT, it does not explain why the band gap decreases with y at a constant Q , as observed. If there is a linear relationship between E_g and Q , it will be approximately $E_g = 2.2 - 2.05Q$, using the band gap of PT (2.2 eV, $Q = 0$) and the E_g value of 0.15 eV calculated for $Q = 1$ by Bredas¹⁹. The expected E_g value of 1.18 eV at $Q = 1/2$ is close to the 1.1 eV found for $y = 1$, but the other values of the band gap at $Q = 1/2$ cannot be explained by this linear equation at all. This is due to the degeneracy of the case $Q = 1/2$. Both the generally small band gaps in this class of polymers and the decrease of E_g with y can be explained in terms of the more primitive concept, the bond-length alternation Δr . The decrease of E_g with y is consistent with the decrease in Δr with expansion of the polymer repeating unit. From the theoretical calculations of Grant and Batra¹⁶ for *trans*-PA and those of Bredas¹⁹ for PT, a linear relationship of the form $E_g = E_g^0 + b\Delta r$ appears to hold. We have obtained the same value of $b = 11.50 \text{ eV/\AA}$ from the two different calculations^{16,19}; however, $E_g^0 = 0$ for *trans*-PA¹⁶ and $E_g^0 = 0.705 \text{ eV}$ for PT¹⁹. This linear equation and the observed band gaps give estimated bond alternations of 0.034, 0.011 and 0.004 Å for $y = 1, 2$ and 3 , respectively, compared to 0.13 Å for PT. It will be interesting to compare these estimated bond alternation values with those

computed or measured directly. Even from these estimates it is clear that nearly uniform bond lengths are attained at $y = 3$ and yet the band gap is as high as 0.75 eV due primarily to E_g^0 . Existence of a finite E_g^0 further distinguishes aromatic and hetero-aromatic conjugated polymers from polyenes¹⁹, and raises the question of its origin, as only the amount above E_g^0 can be attributed to bond alternation. Narrower band gaps would imply negative bond alternation ($\Delta r < 0$), in which carbon-carbon double bonds are longer than single bonds. In principle, as the calculations of Bredas suggest¹⁹, the proposed hetero-aromatic polymers (Fig. 1c) with $Q \rightarrow 1$ could exhibit band gaps smaller than E_g^0 , but such members of this class of polymers have yet to be synthesized.

Obviously, other hetero-aromatic or aromatic polymers with structures similar to Fig. 1c, and hence similar properties, can be expected when the hetero-aromatic units, $(C_4H_2X)_{y,x}$, are replaced by phenyl (Ph), *p*-phenylene oligomers, or other groups capable of both aromatic and quinonoid bonding structures. After we had already synthesized the polymers described here we became aware of the theoretical work of others, predicting a narrow band gap for a structurally related hypothetical (not yet synthesized) polymer. Boudreaux *et al.*²⁴ have calculated a band gap of 1.17 eV and predicted the existence of soliton excitations in poly-*p*-phenylenemethine, $(-Ph-CH=Ph-CH-)_m$.

This work has been partly supported by the US Office of Naval Research. The technical assistance of Marcia Hansen is appreciated.

Received 31 March; accepted 4 June 1986.

- Shirakawa, H., Louis, E. J., MacDiarmid, A. G., Chiang, C. K. & Heeger, A. J. *JCS chem. Commun.*, 578 (1977).
- Chiang, C. K. *et al. Phys. Rev. Lett.* **39**, 1098–1101 (1977).
- Chien, J. C. W. *Polyacetylene: Chemistry, Physics and Materials Science* (Academic, Orlando, 1984).
- Ivory, D. M. *et al. J. chem. Phys.* **71**, 1506–1507 (1979).
- Clarke, T. C. *et al. J. Polym. Sci. Polym. Phys.* **20**, 117–130 (1982).
- Diaz, A. F., Kanazawa, K. K. & Gardini, G. P. *JCS chem. Commun.*, 635 (1979).
- Kanazawa, K. K. *et al. Synth. Metals* **1**, 329–336 (1980).
- Tourillon, G. & Garnier, F. J. *electroanal. Chem.* **135**, 173–178 (1982); **161**, 407–414 (1984).
- Wellinghoff, S. T., Kedrowski, T., Jenekhe, S. A. & Ishida, H. *J. Phys. Colloq.* **44**, 677–681 (1983).
- Jenekhe, S. A., Wellinghoff, S. T. & Deng, Z. *Synth. Metals* **10**, 281–292 (1985).
- Jenekhe, S. A., Wellinghoff, S. T. & Reed, J. F. *Mol. Cryst. Liq. Cryst.* **105**, 175–189 (1984).
- Greene, R. L. & Street, G. B. *Science* **226**, 651–656 (1984).
- Wudl, F., Kobayashi, M., Colaneri, N., Boysel, M. & Heeger, A. J. *Mol. Cryst. Liq. Cryst.* **118**, 199–204 (1985).
- Peierls, R. E. *Quantum Theory of Solids*, 108 (Oxford University Press, London, 1955).
- Ovchinnikov, A. A., Ukrainskii, I. I. & Krentsel, G. V. *Soviet Phys. Usp.* **15**, 575–591 (1973).
- Grant, P. M. & Batra, I. P. *Solid State Commun.* **29**, 225–229 (1979).
- Simon, J. & Andre, J.-J. *Molecular Semiconductors*, 166–173 (Springer, New York, 1984).
- Jenekhe, S. A. *Macromolecules* (submitted).
- Bredas, J. L. *Mol. Cryst. Liq. Cryst.* **118**, 49–56 (1985).
- Su, W. P., Schrieffer, J. R. & Heeger, A. J. *Phys. Rev. B* **22**, 2099–2111 (1980); *Phys. Rev. Lett.* **42**, 1698–1701 (1979).
- Bredas, J. L., Themans, B., Andre, J. M., Chance, R. R. & Silbey, R. *Synth. Metals* **9**, 265–274 (1984).
- Bredas, J. L., Chance, R. R. & Silbey, R. *Phys. Rev. B* **26**, 5843–5854 (1982).
- Jenekhe, S. A. *Macromolecules* (submitted).
- Boudreaux, D. S. *et al. Phys. Rev. B* **31**, 652–655 (1985).

Extraordinary effects of mortar-and-pestle grinding on microstructure of sintered alumina gel

W. A. Yarbrough & Rustum Roy

Materials Research Laboratory, The Pennsylvania State University, University Park, Pennsylvania, USA

The use of the mortar and pestle in the laboratory dates back to the earliest attempts to understand and use the materials around us. Even now, it is difficult to imagine an alternative to crushing and grinding for the preparation of many samples. Traditionally, mortar-and-pestle mixing or grinding has been held to be a relatively mild and controlled process, although it was recognized that

the grinding of materials harder than the material of the mortar and pestle would result in some contamination. Thus, good practice forbade grinding of, for example, mullite and alumina in agate mortars. That this technique is not so innocuous was demonstrated by Dachille and Roy¹ with respect to the stresses generated. They showed that phases usually obtained only at 10–20 kbar can be obtained metastably by simple grinding in a laboratory mortar. There are numerous examples of solid-state phase transformations being produced by prolonged or intense comminution², two of the best-documented cases of stress/shear-induced transformation being the conversion of calcite to aragonite¹, and of quartz to amorphous silica^{3,4}, both by prolonged grinding in a laboratory mortar and pestle. Here we show that the very small amounts of material generated by the wear of mortar and pestle surfaces by even mild grinding can also have substantial effects on the microstructure and transformation kinetics of certain ceramic systems so treated.

Materials prepared by the sol-gel method are typically obtained as non-crystalline or poorly crystallized metastable solids which may then be calcined for densification and/or transformation to a more stable crystalline phase. In the preparation of dense non-crystalline solids (glasses), uncontrolled crystallization or devitrification is undesirable, as this usually leads to a loss of mechanical integrity and other desired properties, such as transparency. In other materials, where a given stable crystalline phase may be the desired product, the kinetics of transformation and the microstructure of the material obtained can be controlled by 'seeding' with the desired phase or with chemically dissimilar materials having, or readily transforming to, the desired structure.

Seeding with crystals of the expected or desired phase has long been used as an aid in the study of phase equilibria, and the importance of even minor amounts of a second phase, isostructural with a stable phase of the composition under study, has already been noted^{5,6}. In particular, we have reported the remarkably enhanced transformation kinetics obtained by seeding sol-gel-derived materials^{7–10}. This may well turn out to be a route to a new class of crystalline solids: nanocomposites (di- or multiphasic solid aggregates), where the size range of the monophasic regions is ~1–100 nm. One of the striking properties of a structurally diphasic xerogel (for example, boehmite sol mixed with an α -alumina sol) is the solid-state epitaxy which occurs on heating, with considerably accelerated kinetics. Thus, Fig. 1 (taken from ref. 9) compares the differential thermal analysis (DTA) of a monophasic boehmite sol with that of a diphasic sol of boehmite and various amounts of α -alumina. Seeding with amounts as small as 0.08 wt % results in a marked lowering of the transformation temperature. Moreover, there is a marked change in the microstructure of the alumina formed.

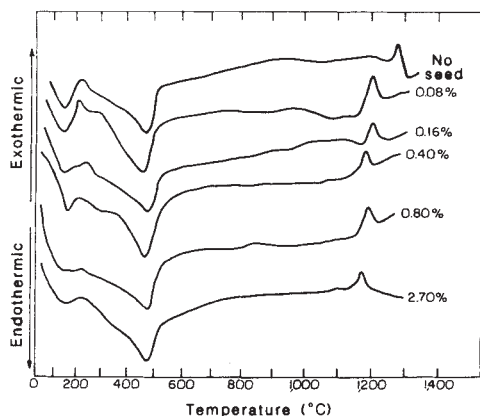


Fig. 1 DTA results showing the effect of seeding with various amounts of α -alumina on the transformation temperature of boehmite gel.

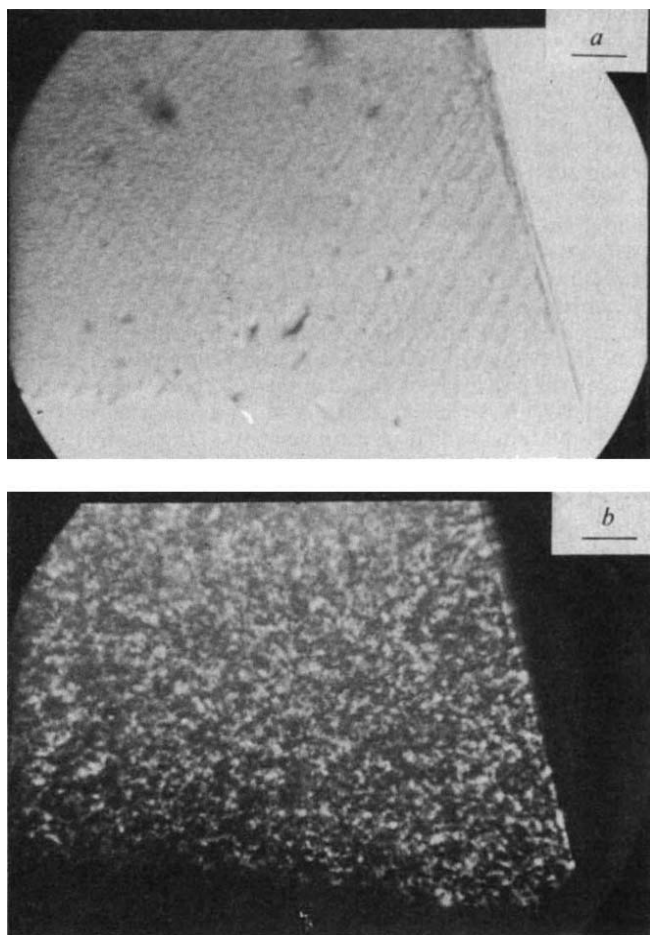


Fig. 2 Boehmite gel fragment prepared from water 'ground' in Diamonite mortar and pestle as seen after calcination at 1,200 °C for 2 h: a, Bright-field transmitted light; b, crossed polars. Scale bars, 10 μ m.

In the unseeded material, monocrystalline cells ~10 μ m across are formed, whereas in the seeded material, the grain size is <1 μ m (ref. 10).

The mortar and pestle experiment was conducted using a Fisher Scientific Co. 'Diamonite' (corundum) mortar and pestle. After thoroughly cleaning the mortar and pestle, 200 ml of deionized water was added to the mortar. The water was slowly 'ground' in the mortar using only light to moderate pressure for 5 min. This water, slightly turbid with suspended detritus from the wear of the mortar and pestle surfaces, was then used to prepare a boehmite (aluminium oxide monohydrate) sol using 13.90 g of a commercially available microcrystalline boehmite. (Dispersal, Remet Chemical Co.) The boehmite was peptized using 15.0 ml of 1.00 M HNO_3 ; the sol was then dried at 90–95 °C to produce a boehmite xerogel. A second batch of xerogel was prepared in precisely the same manner using the same lots of material (water, acid and boehmite), the only modification being the use of a mortar and pestle of agate instead of Diamonite. Fragments of these xerogels were lightly crushed in an agate mortar and pestle. Calcination was performed in air using a Lindberg model 51524 furnace equipped with a model 59256-E1 controller (Lindberg Co.). This furnace uses exposed MoSi_2 elements and had been modified by installing a type S thermocouple positioned just above the sample for accurate temperature measurement. The maximum calcination temperature was 1,200 °C for a period of 2 h, which is sufficient to convert all of the material to α -alumina. The powders were dispersed in immersion oil and examined with a polarizing microscope.

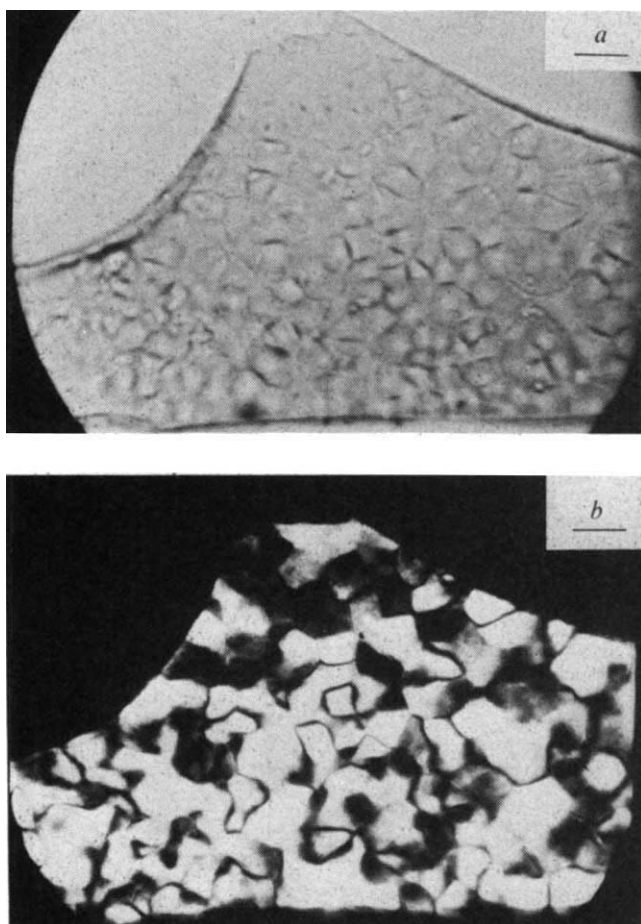


Fig. 3 Boehmite gel fragment prepared from water 'ground' in agate mortar and pestle as seen after calcination at 1,200 °C for 2 h: *a*, Bright-field transmitted light; *b*, crossed polars. Scale bars, 10 μm .

Alpha aluminium oxide (α -alumina) is anisotropic and uniaxial negative. Assuming that a sufficient crystallite size is attained on transformation, its birefringence makes observation in a polarizing microscope relatively easy. Figures 2 and 3 show transmitted-light photomicrographs (Figs 2*a* and 3*a* in bright-field transmitted light; Figs 2*b* and 3*b* using crossed polars) of the calcined powders. Transformation occurs by the nucleation and growth of the α phase in a matrix of very finely divided and poorly crystallized transitional aluminas (δ and θ). The microstructure exhibited in Fig. 2 is identical to that obtained in a xerogel seeded with α -alumina or with materials having the corundum crystal structure; that seen in Fig. 3 is identical to that seen with unseeded or non-isostructurally seeded gels.

Scientists almost universally use milling or grinding in a mortar and pestle as a means of preparing powdered materials, in industry as well as in the laboratory. It is so routine that in many cases no report is made of the materials and methods used¹¹. Enhanced transformation kinetics and nucleation frequency in milled materials have been variously attributed to particle size effects¹² and possibly lattice strain¹³. Dynys and Halloran¹³, in fact, considered and rejected the possibility that detritus from milling was responsible for these effects. In their experiment, α -alumina milling debris was collected, dried and dry-mixed (at 1 wt%) with powders obtained from the calcination of ammonium alum. Pressed pellets were prepared and fired in air. No effect could be seen either by transmission electron microscopy or by comparison of the transformation

kinetics of doped and undoped material. Assuming that the nature of the interface between seed and bulk phase, and the uniformity of seed distribution, are of controlling importance, it is not surprising to find that the efficacy of seeding depends on how the seed phase is produced and introduced to the bulk material. Effects due to the method of seed introduction have been observed previously¹⁴. This demonstration of the power of trace contaminants introduced by grinding has led us to question our common laboratory and commercial practice.

This work was supported by a grant from the Air Force Office of Scientific Research, AFOSR grant no. 83-0212.

Received 17 March; accepted 24 April 1986.

1. Dacheville, F. & Roy, R. *Nature* **186**, 34-71 (1960).
2. Lin, I. J., Nadv, S. & Grodzian, D. J. M. *Miner. Sci. Engng* **7**, 313-336 (1975).
3. Ray, R. C. *Proc. R. Soc. A* **102**, 640-642 (1923).
4. Dale, A. J. *Trans. Br. ceram. Soc.* **23**, 211-216 (1923).
5. Ervin, G. & Osborn, E. F. *J. Geol.* **59**, 331-384 (1951).
6. Roy, D. M. *Am. Miner.* **39**, 140-143 (1954).
7. Suwa, Y., Roy, R. & Komarneni, S. *J. Am. ceram. Soc.* (in the press).
8. Roy, R., Suwa, Y. & Komarneni, S. in *Ultrastructure Processing of Ceramics, Glasses and Composites* (eds Hench, L. L. & Ulrich, D. R.) (Wiley, New York, 1984).
9. Suwa, Y., Komarneni, S. & Roy, R. *J. Mater. Sci. Lett.* **5**, 21-24 (1986).
10. Yarbrough, W. A. & Roy, R. *J. Mater. Res.* (submitted).
11. Guglielmi, M. & Principi, G. *J. Non-Cryst. Solids* **48**, 161-175 (1982).
12. Gregory, A. G. & Veasey, T. J. *J. Mater. Sci.* **7**, 324-332 (1973).
13. Dynys, F. W. & Halloran, J. W. *J. Am. ceram. Soc.* **65**, 442-448 (1982).
14. Bykhovskii, A. I. *Solid State Transformations* (eds Sirota, N. N., Gorskii, F. K. & Varikash, V. M.) (Institute of Solid State Physics and Semiconductors, Academy of Sciences of the Byelorussian SSR, Minsk; Transl. by Consultants Bureau, New York, 1966).

Crustal structure of the northern Alpha Ridge beneath the Arctic Ocean

D. A. Forsyth*, I. Asudeh*, A. G. Green* & H. R. Jackson†

* Lithosphere and Canadian Shield Division, Geological Survey of Canada, 1 Observatory Crescent, Ottawa, Ontario, Canada K1A 0Y3

† Atlantic Geoscience Centre, Geological Survey of Canada, Bedford Institute of Oceanography, PO Box 1006, Dartmouth, Nova Scotia, Canada B2Y 4A2

Beneath the Arctic Ocean, the Alpha Ridge complex is the principal tectonic feature between the Canada Basin and the Lomonsov Ridge. Here we report on analysis of seismic refraction data along the Alpha Ridge and within the basin area 175 km to the north (Fig. 1), which, together with geological and aeromagnetic constraints, indicates that the Alpha Ridge is characterized by regions of indentational-like thickness but has probably resulted from an oceanic mode of development. The ridge flank thins towards the North Pole but may extend as far north as 88° N, and is therefore >500 km wide in the region near 110° W. Consequently, south of the pole along 110° W, typical oceanic crust beneath the eastern Makarov Basin may be present only locally, if at all. Similarities between the velocity structures of the Alpha Ridge and the Iceland area suggest that the Icelandic structure may be a currently active analogue for the Alpha Ridge.

The 1983 CESAR (Canadian Expedition to Study the Alpha Ridge) crustal refraction survey consisted of a 200-km 'strike' line with 7-8-km station spacing along the CESAR South Ridge (C.S.R. in Fig. 1) of the Alpha complex, a parallel 120-km-long 'basin' line north of 87° N, in what we considered to be the Makarov Basin, and a 175-km 'dip' line connecting the other two lines. In addition, an 80-km 'camp' line with an effective spacing of 1-2 km was recorded along the CESAR North Ridge (C.N.R. in Fig. 1). All lines were effectively reversed; the strike line was also recorded to each end from a central shotpoint. The camp and basin lines were shot along the line and recorded at the ends. Multi-element arrays¹ on the ice surface recorded

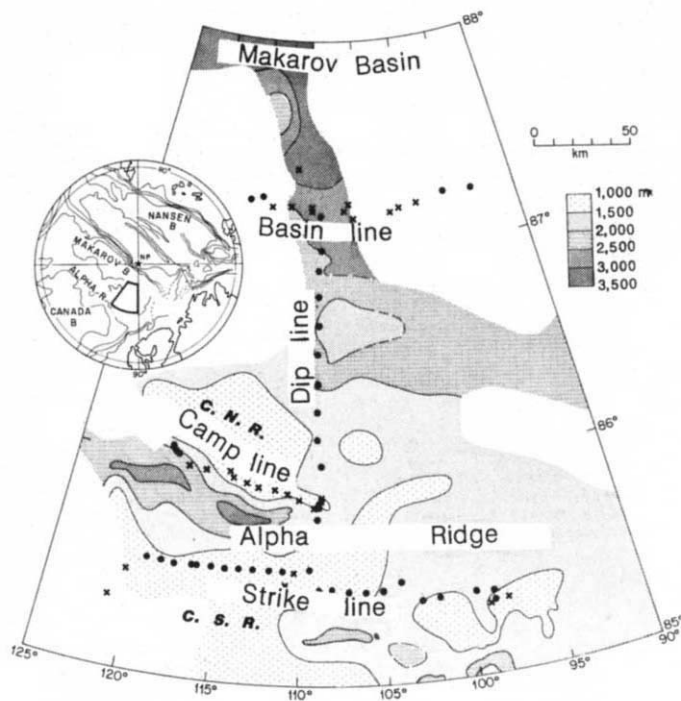


Fig. 1 Location of the CESAR refraction profiles. ●, Recorder location; ×, shot location.

energy from 100–500-kg shots detonated at depths of 100 m (end shots) to 200 m (line shots). We report here the crustal models determined from the strike and basin line data sets. Interpretations of the dip and camp line data are consistent with these models (I.A., A.G.G. and D.A.F., in preparation).

As in most marine surveys, reverberations in the water and sediment layers are common in the CESAR data, and secondary energy arriving later but with apparent P-wave velocity similar to that of the first arrivals can be traced up to several seconds after the first arrivals (Fig. 2A). A thorough study of the nature of these multiples using synthetic seismogram modelling techniques produced water depth estimates in good agreement with known values, and determined that sediment cover was generally <1 km over the ridge area. Sediment thicknesses varying from zero to 1 km were observed independently from air-gun reflection data gathered at the main camp on the ice as it drifted from the central region of the strike line to C.N.R.² (Fig. 1).

The first-arrival waveform of the strike line data remains remarkably similar and continuous for distances up to ~100 km. Within this range, the first-arrival waveforms recorded from the central and eastern shotpoints are similar to the data section in Fig. 2A. There is little evidence of reflected arrivals from mid- or upper-crustal boundaries; the data suggest considerable lateral homogeneity in the upper crust. At distances >100 km the first-arrival waveform uniformity is gradually lost to the effects of deeper lateral velocity variations and reflected arrivals from the crust–mantle boundary. The maximum reflected energy is concentrated in the distance range 150–180 km. The wide-angle reflected energy and the cross-over to first-arrival velocities of ~8.1 km s⁻¹ in the distance range 180–200 km require a relatively thick crustal section.

In Fig. 2A, the synthetic section generated³ from the western shotpoint through our proposed crustal model is compared to the observed data. The strike line model results from synthetic seismogram modelling of the sections from each end of the line and the data recorded from the centre shotpoint. Overlapping and reversed ray paths demonstrate that the model is relatively

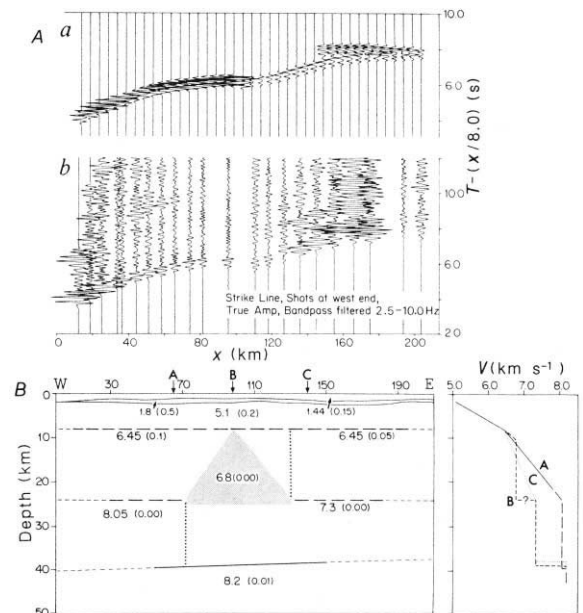


Fig. 2 CESAR strike line. A, comparison of section shot from west end (b) with first-arrival and PmP reflection synthetic section (a). The data (b) are shown with true amplitudes, after application of a bandpass filter (2.5–10.0 Hz). B, Velocity model derived from analysis of reversed strike line data. Velocities (km s⁻¹) are followed by gradients (s⁻¹), in parentheses. Note the central region of relatively lower average velocity.

well constrained to a depth of ~24 km. The exact shape of the anomalous triangular zone of lower average velocity, the velocity structure of the lower crust and the depth of the crust–mantle transition are more weakly constrained. In Fig. 2B the dashed and dotted lines represent velocity or velocity gradient changes (heavy dashes indicate regions constrained by ray-trace modelling). The crust–mantle transition is shown as a solid line to represent the velocity step required to produce the wide-angle reflections in Fig. 2A.

Figure 2B includes the time delays required by variations in water depth and by our best estimates of sediment cover thickness. Beneath the sediments, the upper crust is characterized by a laterally uniform velocity of 5.1 km s⁻¹ and a gradient of ~0.23 s⁻¹. At 9 km depth the velocity gradient diminishes to 0.05 s⁻¹ and 0.10 s⁻¹, in the eastern and western regions respectively, to satisfy times and amplitudes of first-arrival energy at distances <90 km. These gradients lead to mid-crustal velocities of 7.3 km s⁻¹ in the eastern part of the model, and upper-mantle-type velocities of near 8.0 km s⁻¹ at similar depths in the west. The anomalous central region beneath the topographic high (Fig. 1), with a near-constant velocity of 6.8 km s⁻¹, is required to match times and amplitudes of first arrivals beyond 100 km. The depth of the crust–mantle transition cannot be determined confidently from first-arrival data because reversed upper mantle arrivals beyond a distance of 190 km were not well defined. However, given the control on mid- to upper-crustal velocity structure and the position and amplitude of first arrivals and wide-angle reflections (Fig. 2A), our modelling indicates a crust–mantle transition at 38–40 km depth beneath the central and eastern regions and mantle-type velocities at 24 km depth beneath the western region. Tests using a range of plausible crustal and upper mantle velocities suggest a possible range of crustal thicknesses of 36–44 km for the eastern and central regions.

Reduced to a water depth of 2.5 km, reversed basin line sections to distances up to 60 km show a continuity of waveform remarkably similar to the strike line data (compare Figs 2A and

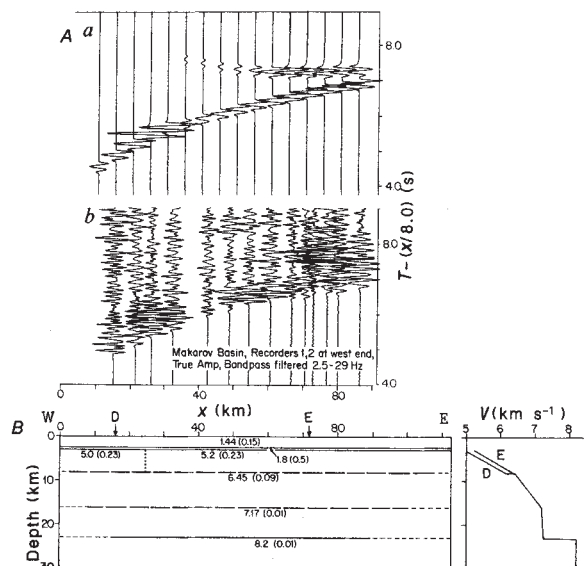


Fig. 3 CESAR basin line. A, comparison of section recorded at west end (b) with first-arrival and PmP reflection synthetic section (a). The data (b) are shown with true amplitudes, after application of a bandpass filter (2.5–29 Hz). B, Velocity model derived from analysis of reversed data obtained at each end of the line.

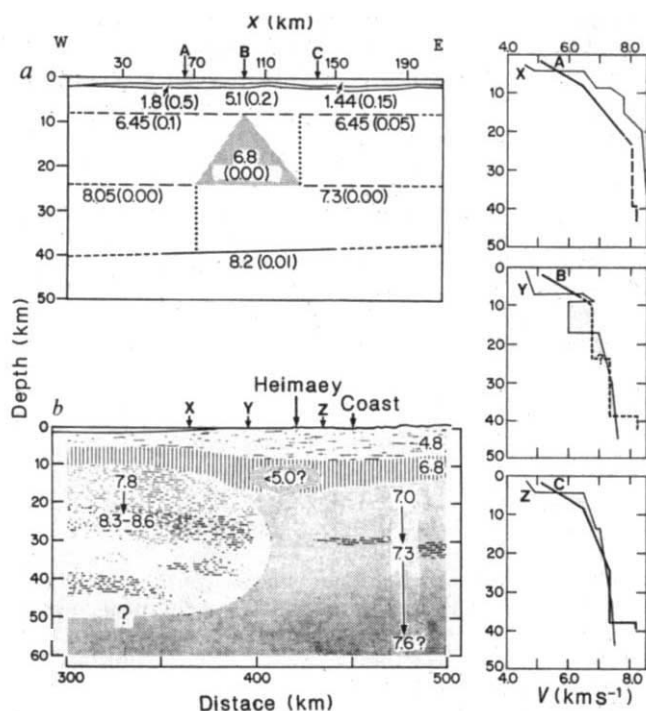


Fig. 4 Comparison of strike line velocity model (a) with model of Iceland-Reykjanes Ridge⁸ (b). Velocity-depth profiles A and X, B and Y and C and Z are compared at right.

3A). Beneath the sediments both data sets require a velocity of 5.1–5.2 km s^{-1} and a gradient of $\sim 0.23 \text{ s}^{-1}$. Thus, the uppermost 15 km of the basin model (Fig. 3B) appear laterally homogeneous and closely resembles the western region of the strike line.

Features observed on the basin line but not observed on the strike line include indications of reflected energy closely following the first arrivals at $\sim 30 \text{ km}$ distance, and a large amplitude secondary event starting at $\sim 60 \text{ km}$. The nearer source reflections result from the larger velocity difference in the western part of the basin (Fig. 3B). Note that the upper-crustal velocity gradient is uniform across the basin. We interpret the high-

amplitude secondary events starting at 60 km as reflections from the crust-mantle transition, similar to those observed on the strike line beyond 120 km. Again, the depth to the crust-mantle transition is poorly constrained, with our best estimate of 23 km being the mean of a range from 21 to 25 km.

Taylor *et al.*⁴ point out that the short-wavelength magnetic anomalies in the area of Fig. 1 closely follow the bedrock topography of the Alpha Ridge, and they propose that ridge material extends farther south beneath the sediments of the Canada Basin than is indicated by the bathymetry. Similarly, the basin line, with a crustal thickness of $\sim 23 \text{ km}$ and a velocity structure similar to the Alpha Ridge, suggests that the ridge complex extends to $\sim 200 \text{ km}$ north of the strike line. These results support the suggestion⁵ that the Alpha Ridge complex extends almost to the edge of the Lomonosov Ridge, and is therefore $> 500 \text{ km}$ wide in this region.

The upper-crustal velocity of 5.1 km s^{-1} immediately underlying a thin sediment cover, the apparent lateral homogeneity along the strike of the ridge, the high-intensity magnetic anomalies⁴ and the vesicular, highly altered volcanic samples dredged from an Alpha Ridge scarp⁶ favour the identification of the upper crustal medium as oceanic layer 2. The greater lateral variation at mid- and lower-crustal depths seems to reflect a more complex tectonic evolution.

The continental-type thickness indicates that the Alpha Ridge, like the Ontong-Java and Manihiki plateaus of the Pacific plate⁷, is an oceanic plateau of problematic origin. Jackson² has outlined the similarity of the Alpha Ridge morphology and shallow structure to that seen on the Manihiki plateau. Certainly, the available geological and geophysical evidence suggests that both the Alpha and Manihiki features are characterized by an over-thickened oceanic-type crust. A closer tectonic analogue to the Alpha structure may be Iceland today. Figure 4 shows a comparison of the velocity structure along the strike of the Alpha Ridge with the structure along the strike of the Mid-Atlantic system from the Reykjanes Ridge to Iceland. The lateral change from a structure characteristic of oceanic crust beneath the flank of the Reykjanes Ridge to a structure with 'anomalous mantle'⁸ beneath Iceland itself resembles the structure along the Alpha Ridge. Note that the crustal structures that have been derived from seismic refraction data may be compared only to depths of $\sim 40 \text{ km}$; Icelandic structure below 40 km is not based on direct refraction data. Comparisons of velocity structures for particular sections of the respective models show:

(1) Thick oceanic sections beneath both the Alpha Ridge and Iceland, with typical layer 2 and layer 3 velocities and a similar velocity-depth profile for the thickened 'anomalous crust' beneath the active area of Iceland and the region beneath the eastern end of the Alpha Ridge profile (Z and C in Fig. 4). The greatest difference between the crustal responses is the interpreted wide-angle reflection coda that indicate significant energy return from a crust-mantle transition near 40 km depth beneath the Alpha Ridge. The absence of similar energy return from beneath Iceland is probably related to the present plume and spreading activity. Comparable activity would have ceased in the Alpha Ridge area during late Mesozoic times².

(2) A central zone (Y and B in Fig. 4), characterized in the Icelandic case by low-velocity regions representing possible magma chambers, and in the Alpha Ridge case by a region of lower velocity gradient. Neither Icelandic nor Alpha Ridge data sets can further resolve the zone of required time delay. In the Icelandic case the lower-velocity regions have been explained as zones of high melt concentration⁸. In the older Alpha Ridge structure, the zone of lower velocity gradient may represent an ancient melt zone which has cooled to yield a higher velocity than in the much younger, still active zone beneath Iceland.

(3) A more normal oceanic velocity-depth profile for the flank of the Reykjanes Ridge and western part of the Alpha Ridge (X and A in Fig. 4), with upper-mantle velocities appearing in the 20–23 km depth range.

Note that, in general, comparable velocities are reached at ~3 km shallower depths on the Icelandic sections than on the Alpha Ridge sections. Although technique resolution may be debated, the results are consistent with the present elevation difference between Iceland (+1.5 km) and the Alpha Ridge (-1.5 km). It is suggested that most of the differences between the velocity-depth profiles may be accounted for by normal thermal decay and subsidence.

We greatly appreciate the valuable discussions and constructive critiques offered by Drs P. Morel-à-l'Huissier, I. Reid, J. R. Weber and M. J. Berry. This is Geological Survey of Canada contribution no. 11086, CESAR contribution no. 7.

Received 10 January; accepted 21 April 1986.

1. Mair, J. A. & Lyons, J. A. *Can. J. Earth Sci.* **18**, 724-741 (1981).
2. Jackson, H. R. *Geol. Surv. Pap. Can.* **84-22** 19-23 (1985).
3. Spence, G. D., Whittall, K. P. & Clowes, R. M. *Bull. seism. Soc. Am.* **74**, 1209-1223 (1984).
4. Taylor, P. T., Kovacs, L. C., Vogt, P. R. & Johnson, G. L. *J. geophys. Res.* **86**, 6323-6333 (1981).
5. Weber, J. R. & Sweeney, J. F. *J. geophys. Res.* **90**, 663-677 (1985).
6. Van Wagoner, N. A. & Robinson, P. T. *Geol. Surv. Pap. Can.* **84-22**, 47-57 (1985).
7. Carlson, R. L., Christensen, N. I. & Moore, R. P. *Earth planet. Sci. Lett.* **51**, 171-180 (1980).
8. Gebrande, H., Miller, H. & Einarsson, P. *J. Geophys.* **47**, 239-249 (1980).

Melting of a model chondritic mantle to 20 GPa

Eiji Ohtani*, Takumi Kato† & Hiroshi Sawamoto†

* Department of Earth Sciences, Ehime University, Matsuyama 790, Japan

† Department of Earth Sciences, Nagoya University, Nagoya 464, Japan

Calculations of the thermal evolution of the Earth during accretion suggest that the outer layer may once have been molten, perhaps to a depth of 1,000 km or more^{1,2}. If such global melting did occur, fractionation processes could have produced a chemical stratification in the mantle. Recent technical developments in experimental petrology have made it possible to achieve temperatures in excess of 2,000 °C at a pressure of 20 GPa^{3,4}, thereby permitting studies of multi-component systems under conditions that prevail in the Earth's mantle. We have conducted melting experiments using a composition in the five-component system CaO-FeO-MgO-Al₂O₃-SiO₂ (CFMAS) corresponding to a model chondritic mantle, and find that the liquidus phase changes from olivine to majorite at a pressure between 12 and 15 GPa. The liquid coexisting with majorite at 20 GPa has a peridotitic composition. In addition, because the CaO/Al₂O₃ ratio of the liquidus majorite at 20 GPa is lower than that of the sub-solidus majorite, partial melting and majorite fractionation at the base of the upper mantle could produce a peridotitic liquid with a CaO/Al₂O₃ ratio greater than that of the chondritic starting material, consistent with current views on mantle geochemistry.

Cosmochemical studies (see, for example, refs 5, 6) suggest that the composition of the bulk Earth is close to chondritic in terms of the major lithophile and refractory elements, such as Mg, Si, Al and Ca. The bulk mantle may contain these elements in ratios close to chondritic. Assuming that the core is composed of iron with ~16 wt % light elements, a mass balance calculation⁷ gives a value of ~0.1 for the Fe/(Fe+Mg) (atomic) ratio of the mantle.

We therefore chose as our starting material a model chondritic mantle composition of 50.18 wt% SiO₂, 3.62% Al₂O₃, 7.16% FeO, 36.18% MgO and 2.86% CaO, in which the ratios of Mg, Si, Al and Ca are chondritic⁸, while the iron content is Fe/(Fe+Mg)=0.1. The samples were prepared by mixing the reagents and heating the mixture at 1,150 °C in an atmosphere with controlled oxygen partial pressure. The starting material after

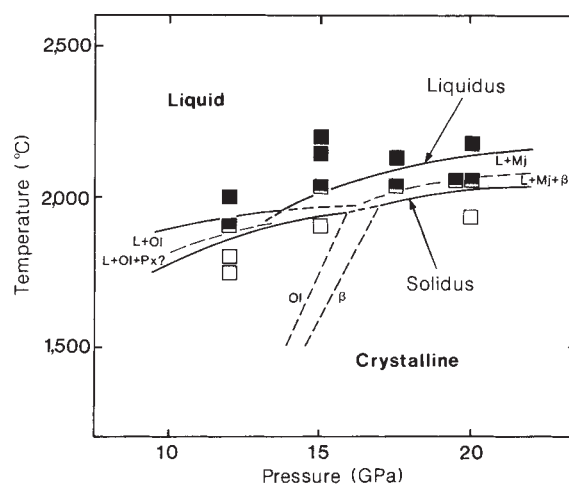


Fig. 1 Melting relations of a model chondritic mantle to 20 GPa. □, Sub-solidus; ■, crystal+liquid; ■, super-liquidus; L, liquid; Ol, olivine; Px, clinopyroxene; β, modified spinel; Mj, majorite.

heating comprised olivine, pyroxenes and a small amount of anorthite.

The apparatus used was an MA8-type high-pressure apparatus³, driven by the guide block/uniaxial press system of Nagoya University, and with a heating assembly similar to that described in ref. 4. The heating element used for the present experiments is a composite of tungsten carbide and diamond. The graphite sample capsule was placed between sheet heaters in the pressure medium, which is made of sintered magnesia containing 8 mol % CoO. A tungsten-rhenium (W25% Re-W3% Re) thermocouple, 0.2 mm in diameter, was placed in contact with the sample capsule, without making electrical contact with the heater. The furnace components and pressure medium were heated to ~1,000 °C for 1 h to remove water. The charge was dried again in an oven at 170 °C for a few hours immediately before the experiments.

The pressure was calibrated using the resistance changes associated with phase changes in Bi, Pb, ZnS and GaAs. A temperature correction was applied to the pressure values calibrated at room temperature, based on the phase boundary curves determined at pressures below 15 GPa by *in situ* X-ray diffraction experiments⁹. The resulting uncertainty in the pressure values was ±0.6 GPa at 15 GPa and ±2 GPa at 20 GPa.

The pressure was applied first, after which the temperature was increased to the desired value and held constant for ~2 min. The charge was quenched by turning off the electrical power to the furnace. The run products recovered after quenching were examined by X-ray powder diffraction, optical and scanning electron microscopy and electron probe microanalyser (EPMA) analysis. The melt is identified under the scanning electron microscope by the presence of fibrous or dendritic texture, which has been established as diagnostic of quenching from liquid at very high pressure^{3,4}.

The conditions and the results of the high-pressure runs are summarized in Fig. 1 as a phase diagram. The run made at 12 GPa and 1,900 °C indicated that olivine is the liquidus phase; thus, the partial melt formed at this pressure is richer in SiO₂ than the starting material with chondritic mantle composition. This result is consistent with melting experiments on the natural peridotite KLB1 (ref. 10). The sub-solidus mineralogy at this pressure is olivine, clinopyroxene and garnet. Figure 2a shows the back-scattered electron image of the charge adjacent to the thermocouple junction, quenched at 20 GPa and 2,050 °C. In this portion of the charge the liquid coexists with two phases, majorite and modified spinel. Figure 2b shows the higher-temperature portion of the same charge, in which the quenched

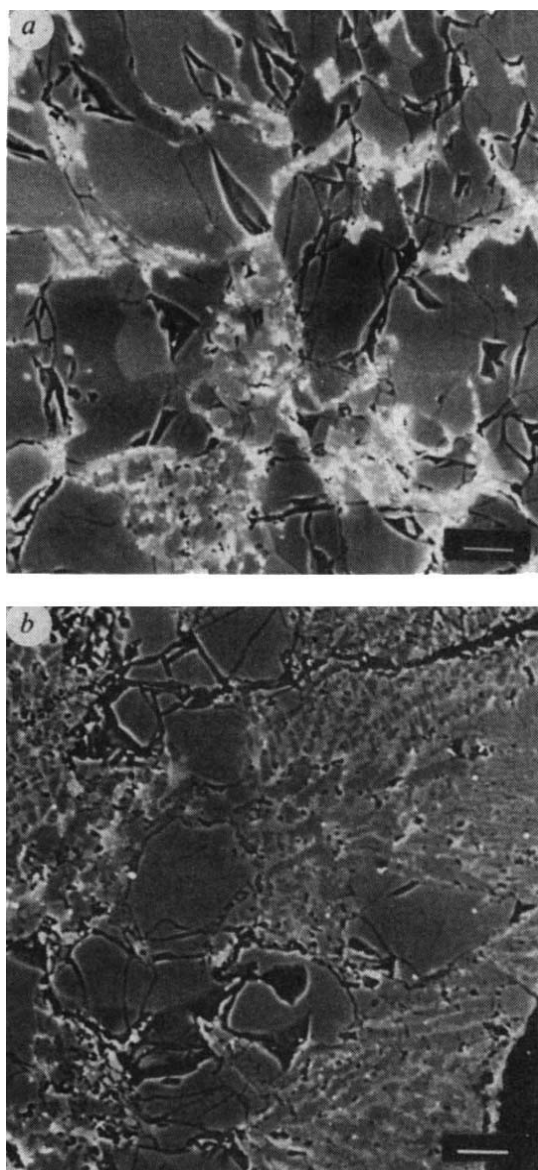


Fig. 2 Back-scattered electron images of the partially molten textures of the model chondritic mantle composition quenched at 20 GPa. In *a*, the quenched liquid (bright parts) is located in interstices between majorite (light grey crystals) and modified spinel (dark grey crystals). In *b*, the quenched liquid, now an aggregate of fine dendritic quench crystals, is coexisting with majorite (large grey crystals). Scale bars, 10 μm .

liquid coexists with granular majorite crystals as a liquidus phase. The temperature difference between the two portions of the charge should have been $<100^\circ\text{C}$, on the basis of a measurement of the temperature distribution in a similar furnace system¹¹. The sub-solidus mineralogy at 20 GPa is a mixture of majorite and modified spinel. Majorite was also observed as the liquidus phase at pressures of 15 and 17.5 GPa. The present experiments clearly demonstrate a change of the liquidus phase from olivine to majorite at a pressure between 12 and 15 GPa for the model chondritic mantle composition.

The compositions of minerals coexisting with a liquid, and of those observed under sub-solidus conditions ($\sim 1,930^\circ\text{C}$), at 20 GPa were determined by EPMA and are given in Table 1. Note that the $\text{CaO}/\text{Al}_2\text{O}_3$ ratio of majorite above the solidus is lower than that observed under sub-solidus conditions; CaO and FeO contents of majorite decrease by partial melting. The partition coefficient K_D of Mg and Fe between majorite and

Table 1 Chemical compositions (wt %) of minerals and partial melt in the model chondritic mantle at 20 GPa

	Ga I	MS I	Ga II	MS II	Ga III	Liq III
SiO_2	56.29	42.89	54.83	42.37	54.24	47.7 ± 2.5
Al_2O_3	4.50	0.29	8.16	1.51	9.70	2.7 ± 1.5
FeO	2.48	4.17	1.88	2.75	1.48	5.1 ± 1.0
MgO	32.60	53.63	34.08	54.22	34.29	40.4 ± 2.5
CaO	3.98	0.09	1.97	0.05	1.14	3.3 ± 1.5
Total	99.85	101.07	100.92	100.90	100.85	99.2
Atomic ratios						
Al/Si	0.094	0.008	0.177	0.042	0.211	
Ca/Al	0.802	0.250	0.218	0.023	0.107	
Fe/Mg	0.043	0.044	0.031	0.028	0.024	

Ga I and MS I, majorite and modified spinel under sub-solidus conditions; Ga II and MS II, majorite and modified spinel coexisting with a liquid (Fig. 2a); Ga III and Liq III, majorite and coexisting liquid.

modified spinel is near unity at 20 GPa and $>1,930^\circ\text{C}$.

The composition of the liquid adjacent to majorite crystals in Fig. 2b was measured by EPMA with a diffused beam, $\sim 15 \mu\text{m}$ in diameter. As the liquid could not be quenched to a glass (quench crystals of modified spinel, majorite and clinopyroxene were formed during quenching), the liquid composition (Table 1) was obtained by averaging measurements obtained over an area of $>50 \times 50 \mu\text{m}$. The liquid is highly magnesian and peridotitic in composition. The FeO contents of the liquid and coexisting minerals shown in Table 1 are lower than would be expected from the starting composition; this is because some of the FeO was reduced to metal during heating at high pressure and temperature. (Small dispersed grains of metallic iron were observed in the charge.) The determination of the liquid composition has a large uncertainty, because of the heterogeneity of the aggregates of quench crystals formed from a liquid during quenching. Nevertheless, our result supports the generation of peridotitic magma by majorite fractionation at the base of a chondritic upper mantle. The low value of $\text{CaO}/\text{Al}_2\text{O}_3$ in the liquidus majorite implies that the partial melt has higher than chondritic $\text{CaO}/\text{Al}_2\text{O}_3$.

The partially molten zone may have extended to the deep upper mantle, or even to the lower mantle in the final stage of the accretion of the Earth^{1,2}. If the zone extended to the base of the upper mantle, the majorite fractionation and upward transportation of the peridotitic liquid could have produced a largely molten, peridotitic upper mantle¹²⁻¹⁴, and a majorite-enriched transition zone. The upper mantle formed by this process would have a higher than chondritic $\text{CaO}/\text{Al}_2\text{O}_3$ ratio, as shown in the present experiment. Thus, the suggestion of Pulme and Nickel¹⁵ that the upper mantle has a high $\text{CaO}/\text{Al}_2\text{O}_3$ ratio, compared with the chondritic value, is consistent with an upper mantle chemistry derived from majorite fractionation by partial melting in the deep upper mantle.

This work was supported by grants 59740186 and 60121007 from the Ministry of Education, Science and Culture, Japan.

Received 26 March; accepted 21 April 1986.

1. Kaula, W. H. *J. geophys. Res.* **84**, 999-1008 (1979).
2. Hayashi, C., Nakazawa, K. & Mizuno, H. *Earth planet. Sci. Lett.* **43**, 22-28 (1979).
3. Ohtani, E. & Kumazawa, M. *Phys. Earth planet. Inter.* **27**, 32-38 (1981).
4. Kato, T. & Kumazawa, M. *Phys. Earth planet. Inter.* **41**, 1-5 (1985).
5. Ganapathy, R. & Anders, E. *Geochim. cosmochim. Acta Suppl.* **5**, 1181-1209 (1974).
6. Anderson, D. L. *J. geophys. Res.* **88**, B41-B52 (1983).
7. Liu, L. G. *Geochim. J.* **16**, 287-310 (1982).
8. Anders, E. & Ebihara, M. *Geochim. cosmochim. Acta* **46**, 2363-2380 (1982).
9. Akimoto, S., Yagi, T. & Inoue, K. in *High-Pressure Research: Applications in Geophysics* (eds Manghnani, M. H. & Akimoto, S.) 595-602 (Academic, New York, 1977).
10. Takahashi, E. & Scarfe, C. M. *Nature* **315**, 566-568 (1985).
11. Irfune, T. & Hiberson, W. O. *High Temp. high Press.* **17**, 575-579 (1985).
12. Herzberg, C. T. & O'Hara, P. J. *Geophys. Res. Lett.* **12**, 541-544 (1985).
13. Walker, D. & Herzberg, C. T. *Eos* **66**, 403-404 (1985).
14. Ohtani, E. *Phys. Earth planet. Inter.* **38**, 70-80 (1985).
15. Pulme, H. & Nickel, K. G. *Geochim. cosmochim. Acta* **49**, 2123-2132 (1985).

Melting of garnet peridotite to 13 GPa and the early history of the upper mantle

Christopher M. Scarfe* & Eiichi Takahashi

Institute for Study of the Earth's Interior, Okayama University, Misasa, Tottori-Ken 682-02, Japan

By analogy with the evolution of the Moon, it has been suggested that the Earth may have had a 'magma ocean' or liquid outer layer before 3,800 Myr ago.¹ Its presence would have had profound implications for the primary stratification of the mantle into upper mantle, transition zone and lower mantle^{2,3}. An essential constraint on this hypothesis is a knowledge of the high-pressure melting characteristics of mantle material, generally assumed to be peridotite in the upper mantle. We recently described⁴ the first melting experiments on mantle peridotite to 14 GPa. Using spinel lherzolite KLB-1 from Kilbourne Hole, New Mexico, we showed that partial melts close to the solidus at 5–7 GPa are komatiitic in composition and that the solidus and liquidus converge at high pressures. Here we report further melting experiments on garnet lherzolite PHN1611 from Thaba Putsoa, Lesotho. We confirm the convergence of the solidus and liquidus and we show a negative dT/dP slope for the liquidus above 7 GPa. These observations are briefly discussed in terms of their importance to the petrological evolution of the upper mantle and the presence of a magma ocean in the early Archaean.

PHN1611 is a sheared garnet peridotite⁵. The bulk composition is given in Table 1 along with analyses of the constituent mineral phases. The composition ($Mg/Mg + Fe = 0.87$; $CaO/Al_2O_3 = 1.19$) is close to pyrolite⁶ and is thought to represent undepleted or fertile mantle lherzolite⁷. The melting relationships of PHN 1611 at pressures ≤ 3.5 GPa have been previously reported, both for the anhydrous rock^{8–10} and in the presence of H_2O (ref. 9) and CO_2 (ref. 11).

Experimental methods and calibration procedures are similar to those in our previous report and are fully described elsewhere¹². The major technological innovation was the use of the 5,000-ton uniaxial, split-sphere, 6–8-type, multi-anvil apparatus developed by Ito^{13,14}. The apparatus incorporates a large sample volume and is capable of pressures of at least 30 GPa and temperatures $> 2,000^\circ C$. Our experiments are the first melting studies on natural rock material using the split-sphere apparatus.

Figure 1 shows that at 1 atmosphere the temperature interval between the liquidus and solidus of PHN1611 is $500^\circ C$: the solidus is at $1,125^\circ C$, the liquidus at $\sim 1,625^\circ C$. With increasing pressure, however, at 5 GPa the solidus is at $1,700^\circ C$ and the liquidus at $1,975^\circ C$. At still higher pressures the solidus increases more gradually, to $1,825^\circ C$ at 13 GPa. In contrast, at pressures > 7 GPa the liquidus temperature no longer increases, but decreases to $1,900^\circ C$ at 13 GPa. The negative slope of the liquidus allows the liquidus and solidus to converge, until at 13 GPa the melting interval is $< 100^\circ C$. In our previous study of KLB-1, results above 10 GPa suggested convergence of the solidus and liquidus, but the results were not definitive enough to confirm convergence⁴.

New determinations of the melting curve of diopside extend the results of Boyd and England¹⁵ from 5 to 13 GPa. Figure 2 shows the melting curve of diopside, along with the melting curves for forsterite¹⁶ and pyrope¹⁷. The diopside melting curve flattens at pressures between 7 and 13 GPa and varies little from $2,000^\circ C$. This behaviour contrasts with that of the melting curve for forsterite, which approximates a straight line with a slope of $dT/dP = 41^\circ C GPa^{-1}$ (ref. 16). However, the curvature of the diopside curve is similar to the strong curvature shown by

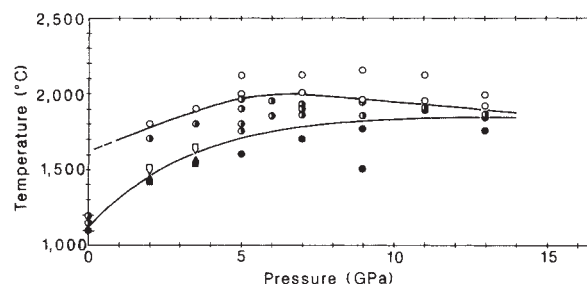


Fig. 1 Phase diagram showing the melting of garnet peridotite PHN1611 to 13 GPa. Experiments close to the liquidus were typically 1 min in duration; however, experiments close to the solidus and in the sub-solidus ranged from 10 to 150 min. Uncertainties in pressure and temperature are discussed in detail elsewhere¹². Because of possible hydration problems, all sample and pressure assemblies were heated with a propane torch before final assembly and were stored in an oven at $120^\circ C$. $\circ$, Super-liquidus; $\bullet$, crystals + liquid; $\bullet$, sub-solidus. Open and filled arrowheads show solidus brackets from refs 8–10.

the melting curve for pyrope, which is attributed to a possible change in Al^{3+} in the melt from 4- to 6-coordination with oxygen¹⁷. Preliminary results by Ohtani and Irifune¹⁸ for diopside to 10 GPa also suggest a flattening of the melting curve at pressures > 8 GPa. At 7–13 GPa, diopside melt does not quench to a glass but exhibits prismatic quench crystals. Note, on the other hand, that at ≤ 5 GPa diopside melt can be quenched to a glass⁵. This could imply a change of melt structure above and below 7 GPa.

It is possible that the negative slope of the peridotite liquidus above 7 GPa is due to the solution of CO_2 or CO in the melt as a result of reduction of iron oxides in the charge by graphite from the graphite capsule. Carbon dioxide reduces melting temperatures at lower pressures^{11,19}; however, both the solidus and liquidus should be affected. Microprobe analyses of carbon in two charges (CMS 62, 13 GPa, $1,920^\circ C$; CMS 52, 11 GPa, $1,900^\circ C$) by T. Furuta (Ocean Research Institute, University of Tokyo) show concentrations of < 0.5 wt % using Al metal as a conductive coating. At this stage, however, these analyses must be considered preliminary because no special precautions were used to exclude carbon during polishing of the charges and because the analyses involved the development of special techniques.

The compositions of melts and coexisting crystals have been reported for spinel lherzolite KLB-1 in our previous paper⁴. The results for PHN1611 are very similar and we report here only analyses of the composition of near-solidus olivine, clinopyroxene and garnet as a function of pressure (Table 1). The major point is that above ~ 5 GPa at temperatures $\geq 1,750^\circ C$ the near-solidus assemblage is olivine + clinopyroxene + garnet. Orthopyroxene was not identified and must be excluded at high pressures by solution of the enstatite component in both clinopyroxene and garnet (Fig. 3). When compared with the composition of the starting material, with increasing pressure the Si and $Mg/Mg + Fe$ contents of the garnet increase and the Al content decreases. At the same time, the clinopyroxene becomes depleted in Al and Ca and enriched in Fe. These trends may be compared with previous work on the sub-solidus phase relationships of PHN 1611 (ref. 20), and on pyroxene-garnet solid solution relationships in Mg- and Fe-rich systems²¹ and in the system $CaO-MgO-Al_2O_3-SiO_2$ (CMAS; ref. 22). Note, however, that any differences may be accounted for by the much higher temperatures of the present experiments compared with temperatures of $\leq 1,500^\circ C$ in previous work. Recently, natural garnets from the upper mantle with pyroxene in solid solution have been described by Moore and Gurney²³. These garnets, found as inclusions in diamond in kimberlite, have similar compositions to garnets in the experiments.

* Permanent address: Department of Geology and Institute of Earth and Planetary Physics, University of Alberta, Edmonton, Canada T6G 2E3.

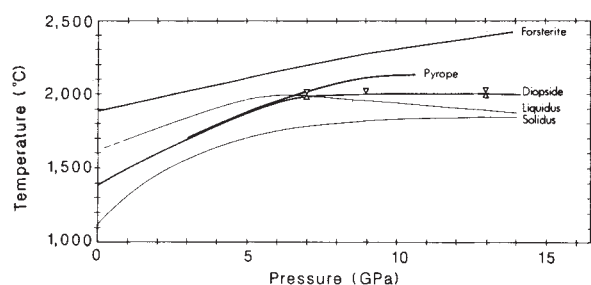


Fig. 2 Melting curves for forsterite, diopside and pyrope, superimposed on the phase diagram for PHN1611. Brackets at high pressures for diopside are from the present study.

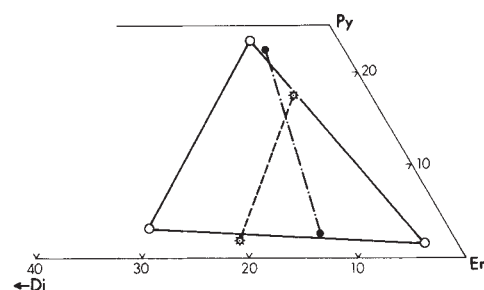


Fig. 3 Garnet-pyroxene chemical relationships projected on to the plane pyrope-enstatite-diopside in CMAS. ○, PHN1611 starting material; ●, charge CMS41 (5 GPa, 1,750 °C); ⊗, charge CMS64 (13 GPa, 1,850 °C).

The present results for garnet lherzolite PHN1611 confirm our previous work on spinel lherzolite and show the convergence of the solidus and liquidus of mantle peridotite at pressures >7 GPa. This result had been anticipated by Herzberg²⁴. Below, we discuss three implications of the results.

First, the question arises as to whether complete convergence takes place between the liquidus and solidus at some pressure above 13 GPa, or whether the curves approach each other but do not actually join (see refs 24–28). In a multi-component system with solid solutions, complete convergence may not occur but a close approach is possible. Therefore, it is anticipated that at a pressure close to 15 GPa the slope of the liquidus will again become positive and a cusp or inflection will be generated where olivine is replaced by garnet-pyroxene solid solution (majorite)^{12,24}. At this cusp, for small temperature excursions across the solidus, melts would be close to the bulk composition of upper mantle peridotite.

Second, because of the negative dT/dP slope of the liquidus, melts in equilibrium with olivine in the pressure interval 7–13 GPa may be denser than olivine, in order to satisfy the Clapeyron relationship. However, note that the Clapeyron relationship is strictly applicable only to systems where the liquid and solid compositions are identical; in a natural multi-

component system where liquids and coexisting crystals have different compositions, additional terms are required^{24,29,30}. Olivine flotation was not confirmed in our experiments, but we believe this may be due to the very short run durations near the liquidus and to the presence of temperature gradients in our charges¹². The possibility that melts at high degrees of partial melting are more dense than olivine has profound implications for melt-crystal dynamics at depth in the Earth. The most obvious conclusion is that melts, once formed at depths >200 km, may not be able to rise to shallower levels, but will remain concealed at depth. The density cross-over between liquid and crystals results from the greater compressibility of silicate melts at high pressure compared with their coexisting crystalline solids^{29,31}. Several petrogenetic possibilities related to this observation have been discussed elsewhere^{12,24–27,29,31,32}.

Third, it has been suggested that the Earth may have possessed a liquid or partially molten 'magma ocean' during the early stages of its evolution^{2,3,26,32}. A similar concept has been discussed in relation to the evolution of the Moon¹. Our present results suggest that because of the close approach of the solidus and liquidus at pressures of ~15 GPa, the Earth could have had a subterranean molten layer of broadly lherzolitic composition. Whether the eutectic-like melting behaviour implies that the

Table 1 Chemistry of garnet peridotite PHN1611 and selected run products

	PHN1611 starting material						CMS41: 5 GPa, 1,750 °C			CMS64: 13 GPa, 1,850 °C		
	PHN1611	Pyrolite	Gt	Cpx	Opx	Ol	Gt	Cpx	Ol	Gt	Cpx	Ol
SiO ₂	44.54	45.20	42.68	54.79	56.26	40.42	42.85	55.78	40.78	46.92	56.03	40.81
TiO ₂	0.25	0.71	0.80	0.30	0.23	0.03	0.24	0.05	0.01	0.51	0.07	0.04
Al ₂ O ₃	2.80	3.54	21.15	2.41	1.34	0.10	21.05	2.58	0.14	15.74	1.57	0.12
FeO(total)	10.24	8.47	8.63	5.14	7.04	11.44	5.78	5.40	8.26	7.29	6.07	11.46
MnO	0.13	0.14	0.26	0.13	0.13	0.14	0.17	0.10	0.12	0.19	0.09	0.10
MgO	37.94	37.48	20.66	20.32	32.66	48.06	23.97	28.25	49.64	23.88	23.91	46.95
CaO	3.32	3.08	4.30	13.03	1.59	0.13	3.96	6.20	0.28	3.72	9.73	0.22
Na ₂ O	0.34	0.57	0.07	1.50	0.34	ND	0.10	0.51	0.05	0.13	0.91	0.06
K ₂ O	0.14	0.13										
P ₂ O ₅	ND	0.06										
Cr ₂ O ₃	0.29	0.43	1.46	0.49	0.21	0.05	1.46	0.37	0.17	1.18	0.27	0.11
NiO	ND	0.20										
Total	99.99	100.01	100.01	98.11	99.80	100.37	99.58	99.24	99.45	99.56	98.65	99.87
Mg/(Mg+Fe)	0.868	0.887	0.810	0.875	0.892	0.882	0.881	0.903	0.915	0.854	0.876	0.878
0.5(Al+Cr)/ [0.5(Al+Cr)+Fe+Mg+Ca]			0.234	0.032	0.016		0.225	0.026		0.176	0.020	
Ca/[0.5(Al+Cr)+Fe+Mg+Ca]			0.083	0.278	0.030		0.073	0.121		0.072	0.200	

According to recent studies (F. R. Boyd, personal communication), PHN1611 may have been affected by metasomatism immediately before eruption; however, the magnitude of any compositional changes is unclear. Analyses of experimental charges were performed at Misasa with a JEOL JXA-5A microprobe. Conditions for wavelength-dispersive analysis were a 15-kV accelerating voltage, a beam current of 20 nA and counting times of 40–70 s. Data reduction using ZAF corrections. ND, Not determined.

upper mantle itself is a partial melt sweated out of the whole mantle^{26,27,33}, and whether on cooling and crystallization the layer would fractionate olivine by flotation and garnet solid-solution by settling^{2,3,26,27}, thus inducing some mineralogical stratification of the upper mantle, are questions that go beyond the scope of this paper. However, the melting phase relationships of garnet peridotite presented here provide a starting point for modelling the early Earth and the evolution of its upper mantle.

This study was supported by grants 58460042 and 59540525 from the Ministry of Education, Science and Culture of Japan. C.M.S. acknowledges the support and hospitality of ISEI and travel funds from Canadian NSERC grant IC-0214 and the University of Alberta Central Research Fund. Samples of PHN1611 were supplied by Drs F. R. Boyd and P. H. Nixon. We thank Dr E. Ito for advice and use of his apparatus. Comments by Drs D. L. Anderson, F. R. Boyd, C. T. Herzberg, E. Ito, Y. Matsui and D. Walker helped to improve the manuscript.

Received 3 December 1985; accepted 16 April 1986.

1. Warren, P. H. *A. Rev. Earth planet. Sci.* **13**, 201–240 (1985).
2. Anderson, D. L. *J. geophys. Res.* **84**, 6297–6298 (1979).
3. Anderson, D. L. *Science* **223**, 347–355 (1984).
4. Takahashi, E. & Scarfe, C. M. *Nature* **315**, 566–568 (1985).
5. Nixon, P. H. & Boyd, F. R. in *Lesotho Kimberlites* (ed. Nixon, P. H.) (Lesotho National Development Corporation, 1973).

6. Ringwood, A. E. in *Advances in Earth Sciences* (ed. Hurley, P. M.) (Massachusetts Institute of Technology, Cambridge, 1966).
7. Shimizu, N. *Earth planet. Sci. Lett.* **25**, 26–32 (1975).
8. Kushi, I. in *Lesotho Kimberlites* (ed. Nixon, P. H.) (Lesotho National Development Corporation, 1973).
9. Mysen, B. O. & Kushiro, I. *Am. Miner.* **62**, 843–865 (1977).
10. Harrison, W. J. *Yb. Carnegie Instn. Wash.* **78**, 562–568 (1979).
11. Wendlandt, R. F. & Mysen, B. O. *Am. Miner.* **65**, 37–44 (1980).
12. Takahashi, E. *J. geophys. Res.* (in the press).
13. Ito, E. & Yamada, H. *Advances in Earth and Planetary Sciences* Vol. 12 (Reidel, Dordrecht, 1982).
14. Ito, E., Takahashi, E. & Matsui, Y. *Earth planet. Sci. Lett.* **67**, 238–248 (1984).
15. Boyd, F. R. & England, J. L. *J. geophys. Res.* **68**, 311–323 (1963).
16. Ohtani, E. & Kumazawa, M. *Phys. Earth planet. Inter.* **27**, 32–38 (1981).
17. Ohtani, E., Irifune, T. & Fujino, K. *Nature* **294**, 62–64 (1981).
18. Ohtani, E. & Irifune, T. *Seism. Soc. Jap. Abstr. C51*, 243 (1985).
19. Egger, D. H. *Am. J. Sci.* **278**, 305–343 (1978).
20. Akaogi, M. & Akimoto, S. *Phys. Earth planet. Inter.* **19**, 31–51 (1979).
21. Akaogi, M. & Akimoto, S. *Phys. Earth planet. Inter.* **15**, 90–106 (1977).
22. Yamada, H. & Takahashi, E. in *Kimberlites II: The Mantle and Crust–Mantle Relationships* (ed. Kornprobst, J.) (Elsevier, Amsterdam, 1984).
23. Moore, R. O. & Gurney, J. J. *Nature* **318**, 553–555 (1985).
24. Herzberg, C. T. *Phys. Earth planet. Inter.* **32**, 193–202 (1983).
25. Ohtani, E. *Earth planet. Sci. Lett.* **67**, 261–272 (1984).
26. Ohtani, E. *Phys. Earth planet. Inter.* **38**, 70–80 (1985).
27. Herzberg, C. T. & O'Hara, M. J. *Geophys. Res. Lett.* **12**, 541–544 (1985).
28. Walker, D. *Contr. Miner. Petrol.* **92**, 303–307 (1986).
29. Rigden, S. M., Ahrens, T. J. & Stolper, E. M. *Science* **226**, 1071–1074 (1984).
30. Walker, D. & Agee, C. *Eos* **67**, 405 (1986).
31. Stolper, E. M., Walker, D., Hager, B. H. & Hays, J. F. *J. geophys. Res.* **86**, 6261–6271 (1981).
32. Nisbet, E. G. & Walker, D. *Earth planet. Sci. Lett.* **60**, 103–113 (1982).
33. Takahashi, E., Matsui, Y. & Ito, E. *Proc. 10th Symp. Antarctic Meteorites*, 59–60 (National Institute of Polar Research, Tokyo, 1985).

Evidence from the Parana of south Brazil for a continental contribution to Dupal basalts

C. J. Hawkesworth*, M. S. M. Mantovani†, P. N. Taylor‡ & Z. Palacz*

* Department of Earth Sciences, The Open University, Milton Keynes, MK7 6AA, UK

† Departamento de Geofísica, Instituto Astronômico e Geofísico, Universidade de São Paulo, Caixa Postal 30627, 01051 São Paulo SP, Brazil

‡ Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK

It has been suggested^{1,2} that the oceanic upper mantle preserves a large scale isotope anomaly, Dupal², which may be thousands of millions of years old. New Nd-, Pb- and Sr-isotope data on the continental volcanic rocks of the Parana, south Brazil, reveal 'enriched' isotopic ratios, with $^{87}\text{Sr}/^{86}\text{Sr} = 0.705\text{--}0.716$ and $\epsilon_{\text{Nd}} = -2.5$ to -8 , similar to those from other continental flood basalt provinces. Their $^{207}\text{Pb}/^{204}\text{Pb}$ ratios are higher than those of mid-ocean-ridge basalts (MORB) at comparable $^{206}\text{Pb}/^{204}\text{Pb}$, and even though the basalts appear to have been derived from lithospheric sources within the sub-continental mantle, they preserve isotope and trace element ratios similar to those in oceanic basalts with the Dupal signature in the South Atlantic. The implied link between these continental flood basalts and Dupal oceanic volcanics raises the possibility that in some areas the Dupal anomaly marks a comparatively shallow-level feature in the Earth's mantle.

The Parana is one of the largest known continental flood basalt provinces (10^6 km^2). Most age determinations cluster around 130–120 Myr (ref. 3), so that the Parana appears to be contemporaneous with the Etendeka volcanics in Namibia⁴ and the opening of the South Atlantic⁵. The Parana rocks range from tholeiitic basalts to rhyodacites and rhyolites. The more evolved rocks tend to occur at higher stratigraphical levels, and in a large regional study by Bellieni *et al.*⁶ they comprised ~16% of the rocks analysed.

Geochemically the Parana volcanics consist of two apparently discrete groups characterized by different minor and trace ele-

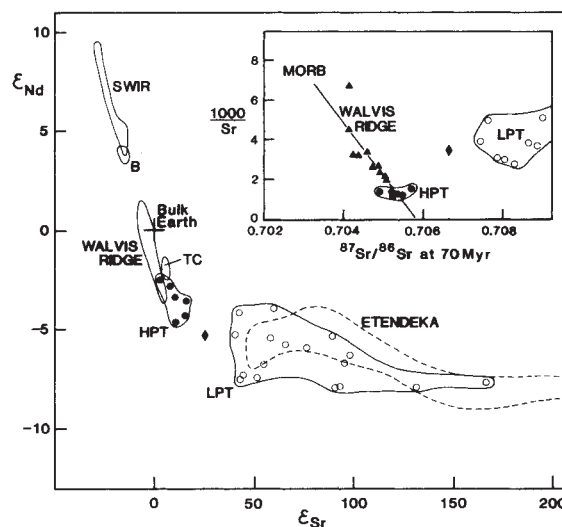


Fig. 1 Plot of ϵ_{Sr} versus ϵ_{Nd} , comparing Parana data with those from: SWIR, south-west Indian Ridge; B, Bouvet; TC, Tristan da Cunha, the Walvis Ridge, and the low-Ti Etendeka rocks from Namibia^{12,22,26,27}. ●, HPT; ○, LPT; ◆, a transitional basalt. Inset, $1,000/\text{Sr}$ versus $^{87}\text{Sr}/^{86}\text{Sr}$ for the Walvis Ridge (▲) and Parana rocks (after ref. 22).

ment abundances and $^{87}\text{Sr}/^{86}\text{Sr}$ ratios^{6–9}. The HPT (high phosphorus and titanium) rocks have high P_2O_5 , TiO_2 , Sr and LREE (light rare-earth element) contents but, at least in the samples analysed here, restricted major element contents ($\text{SiO}_2 = 49.3\text{--}53.9\%$, $\text{MgO} = 5.0\text{--}4.3\%$) and initial $^{87}\text{Sr}/^{86}\text{Sr}$ (0.7048–0.7058). LPT (low phosphorus and titanium) rocks exhibit a full range from basalt to rhyolite, but have low P_2O_5 and TiO_2 (<0.3 and 2.0%, respectively), and high and variable initial $^{87}\text{Sr}/^{86}\text{Sr}$ (0.7076–0.7160). Mantovani *et al.*⁸ explained the striking increase in $^{87}\text{Sr}/^{86}\text{Sr}$ with differentiation in the LPT rocks in terms of crustal assimilation during high-level fractional crystallization (AFC). However, they and others^{6–9} have argued that the HPT and LPT series were derived from sources with different minor and trace element features, and probably different $^{87}\text{Sr}/^{86}\text{Sr}$, of ≈ 0.705 and ≈ 0.707 respectively.

Regional studies of the Parana have demonstrated that, as in

Table 1 Parana volcanics

	SiO*	TiO ₂ *	Rb*	Sr*	Sm*	Nd*	⁸⁷ Sr/ ⁸⁶ Sr†	¹⁴³ Nd/ ¹⁴⁴ Nd†	²⁰⁶ Pb/ ²⁰⁴ Pb†	²⁰⁷ Pb/ ²⁰⁴ Pb†	²⁰⁸ Pb/ ²⁰⁴ Pb†
GB 28ab	49.6	3.87	34	736	11	58	0.70540	0.512437 ± 8	17.558	15.506	38.090
GB 33ab	49.4	3.76	22	740	10	50	0.70499	0.512451 ± 24	17.087	15.445	37.240
GB 23ab	49.5	3.62	54	816	11	53	0.70565	0.512415 ± 14	17.159	15.482	37.659
GB 45ac	49.5	3.85	14	755	11	53	0.70540	0.512351 ± 16	16.629	15.501	38.132
GB 13af	50.9	3.53	22	663	10.5	49.0	0.70581	0.512368 ± 14	17.062	15.447	37.388
GB 41ac	50.8	3.44	43	304	11.6	58.4	0.70705	0.512328 ± 20	17.727	15.529	38.315
GB 57da	48.7	1.43	10	254	4.61	19.9	0.70755	0.512330 ± 14	18.495	15.628	38.621
GB 43ab	48.9	1.86	59	329	5.8	25	0.70856	0.512221 ± 16	18.096	15.593	38.639
GB 20a	49.5	0.80	15	188	2.58	10.6	0.70924	0.512413 ± 26	18.299	15.602	38.566
GB 40bb	49.7	0.90	9	199	3.05	12.9	0.70779	0.512389 ± 24	18.523	15.631	38.655
GB 45ab	50.6	1.75	16	375	6.2	29	0.70846	0.512207 ± 24	17.942	15.583	38.356
GB 50ac	50.9	1.69	41	259	5.8	20	0.70918	0.512281 ± 20	18.358	15.622	38.642
GB 18ac	51.1	0.99	22	228	3.85	16.6	0.70951	0.512302 ± 20	18.068	15.624	38.323
GB 39aa	51.2	1.77	21	271	5.8	22	0.70907	0.512339 ± 8	18.820	15.664	38.797
GB 44ad	51.5	1.72	41	305	5.0	26	0.70824	0.512194 ± 18	17.201	15.497	37.442
GB 2ah	53.1	0.87	35	237	4.73	22.2	0.71187	0.512193 ± 20	18.269	15.617	38.541
GB 6ac	54.2	1.22	50	225	6.61	28.7	0.71196	0.512184 ± 18	18.521	15.633	38.855
GB 36aa	54.5	1.75	72	255	6.10	28.1	0.71220	0.512319 ± 34	18.420	15.651	38.370
GB 53ab	55.0	1.43	73	138	7.2	35	0.71588	0.512177 ± 14	—	—	—
GB 54a	57.6	1.48	54	168	7.82	37.2	0.71769	0.512199 ± 20	19.024	15.705	38.957
H 295	53.2	1.42	38	178	4.6	20	0.71239	0.512258 ± 18	18.687	15.658	38.800
H 205	54.1	1.44	43	196	5.7	22	0.71100	0.512308 ± 16	18.624	15.660	38.926
T 4	53.3	1.29	37	303	5.1	24	0.71213	0.512266 ± 16	18.764	15.682	39.919
T 23	69.5	0.97	139	148	8.5	38	0.72009	0.512202 ± 10	19.114	15.727	38.987

* Data from ref. 8 except for Sm and Nd results with three significant figures, which are by isotope dilution.

† Measured ratios. Analytical errors = 0.007% for ⁸⁷Sr/⁸⁶Sr and 0.11% for Pb. ⁸⁷Sr/⁸⁶Sr in NBS987 = 0.71018 ± 4 and ¹⁴³Nd/¹⁴⁴Nd in BCR-1 = 0.51262 ± 2. Pb isotope ratios corrected for -0.1%/AMU mass fractionation, assessed from repeated analysis of NBS 981.

the Karoo¹⁰⁻¹¹, the northern rocks are predominantly high-Ti, whereas those in the south are low-Ti (ref. 10). Our samples, for which fuller descriptions are given elsewhere⁸, are from close to the probable boundary between high- and low-Ti provinces⁹. Most are from a detailed section at 28°S, in which high-Ti rocks are confined to a 200-m segment (30%) near the base.

All samples have higher ⁸⁷Sr/⁸⁶Sr and lower ¹⁴³Nd/¹⁴⁴Nd ratios than the bulk Earth (Table 1). HPT rocks show restricted ϵ_{Sr} (from 4 to 16) and ϵ_{Nd} (from -2.5 to -4.6), and plot on an extension of the so-called mantle array (Fig. 1). The LPT rocks have slightly lower ϵ_{Nd} (from -3.7 to -7.9), but higher and more variable ϵ_{Sr} , plotting in a flat-lying field on the $\epsilon_{\text{Nd}}-\epsilon_{\text{Sr}}$ diagram similar to that for contemporaneous lavas in Namibia which are also low-Ti (refs 11-13).

Pb isotope data for the Parana define arrays sub-parallel to MORB, but displaced to higher ²⁰⁷Pb/²⁰⁴Pb and ²⁰⁸Pb/²⁰⁴Pb (Fig. 2). The slope of the array in the ²⁰⁷Pb/²⁰⁴Pb-²⁰⁶Pb/²⁰⁴Pb diagram corresponds to an apparent age of ~2,000 Myr, and the array yields a single-stage model μ value of 8.11. The Pb isotopes straddle the first-stage geochron for 120 Myr (the best estimate of eruption age), with the HPT rocks generally having less radiogenic Pb (²⁰⁶Pb/²⁰⁴Pb from 17.06 to 17.73) than the LPT (Fig. 2).

Crustal contamination is a possibility during continental magmatism^{14,15}. For the Parana, the present consensus is that the HPT exhibit no clear signs of significant contamination⁶⁻⁹. They evolved in an environment characterized by high Ti, Ta and Sr abundances, and with trace element ratios similar to OIB (ocean island basalts), presumably in the sub-continental mantle. In contrast, the striking changes in isotope ratios with SiO₂ in the LPT are due to open-system differentiation within the crust, and determination of the isotope ratios of uncontaminated LPT remains problematical. Published models emphasize that primitive HPT and LPT magmas had different isotope and trace element ratios, perhaps with ϵ_{Sr} = 30-40 (and by implication from Pb-Sr correlations, ²⁰⁶Pb/²⁰⁴Pb = 18) in uncontaminated LPT⁶⁻⁹.

On the ²⁰⁷Pb/²⁰⁴Pb-²⁰⁶Pb/²⁰⁴Pb diagram the HPT and LPT

data lie on sub-parallel, slightly en echelon, trends (Fig. 2). The LPT field is dominated by progressive contamination, with more radiogenic Pb in the higher SiO₂ rocks. The HPT rocks, which show no evidence of crustal contamination, plot approximately on the extension of the LPT trend at less radiogenic compositions, and define an apparent isochron age of 1,800 ± 400 Myr (MSWD = 0.7), with a model μ value of 8.1. This age is very similar to a Pb-Pb age of 2,180 ± 180 Myr (model μ = 8.06) recently reported by Mantovani *et al.*¹⁶ for nine composite samples of basement rocks from beneath the Parana, and the simplest interpretation is that the major crust-forming event in this area took place ~2,200 Myr ago, and that in the following 200-300 Myr mantle material stabilized within the continental lithosphere. Such material probably included mantle depleted after crust extraction, and variably enriched by trapped melts and fluids. It was later remobilized in the Parana event, and from the initial isotope ratios of the HPT rocks, it had average U/Pb ≈ 0.08, Sm/Nd ≈ 0.27 and Rb/Sr ≈ 0.049. Comparison with measured trace element ratios in the HPT indicates that in this model Sm/Nd was on average reduced by 18%, and Rb/Sr increased by 8%, during melting and subsequent fractionation.

Coupled crust-mantle domains have been invoked elsewhere^{13,17,18}, but the question is whether material from the continental lithosphere subsequently contributes to the composition of the oceanic basalts^{19,20}.

Hart's recent survey of isotopic data from both oceanic and rift-related continental basalts indicated that, while many scatter around a Pb/Pb array that includes most MORB, there are areas where the basalts are consistently displaced to relatively high ²⁰⁷Pb/²⁰⁴Pb and ²⁰⁸Pb/²⁰⁴Pb² (ref. 2). Termed the Dupal anomaly, one such area is in the South Atlantic (Fig. 3), where rocks with $\Delta 7/4 > 10$ and $\Delta 8/4 > 100$ tend also to have high ⁸⁷Sr/⁸⁶Sr, ($\Delta \text{Sr} > 45$). The relation between Pb and Sr isotopes is complex, and although the relatively high-²⁰⁷Pb/²⁰⁴Pb rocks tend to have higher ⁸⁷Sr/⁸⁶Sr, there is no simple correlation between ⁸⁷Sr/⁸⁶Sr and ²⁰⁶Pb/²⁰⁴Pb along the Pb/Pb arrays.

Variations in ²⁰⁷Pb/²⁰⁴Pb and ²⁰⁸Pb/²⁰⁴Pb require long-lived variations in U/Pb and Th/Pb (>10⁹ yr), but they may be

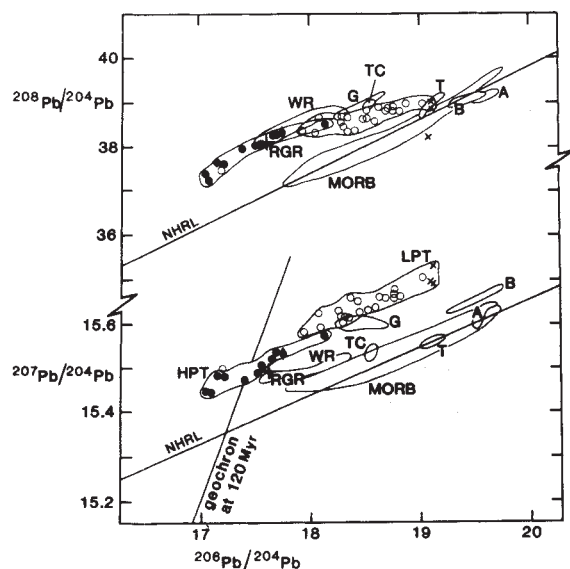


Fig. 2 Pb-isotope diagrams. Parana symbols as in Fig. 1. NHRL, the Northern Hemisphere Reference Line from ref. 3. RGR, Rio Grande Rise; WR, Walvis Ridge; TC, Tristan de Cunha; G, Gough; T, Trinidad; A, Ascension, B, Bouvet^{22,28}.

interpreted as evidence either for old, anomalous regions in the sub-lithospheric mantle, or for the introduction of anomalous Pb from lithospheric reservoirs. Hart² preferred the former because recent injection of crustal material would tend to generate Pb/Pb arrays at a high angle to the MORB trend, rather than sub-parallel to it. Weaver *et al.*²¹ demonstrated that in the South Atlantic, basalts from the area of the Dupal anomaly are characterized by higher La/Nb, Th/Ta, Th/U, Ba/La and Ba/Nb than basalts from outside the anomaly. Such high ratios are relatively unusual in OIB and were attributed by Weaver *et al.* to the introduction of a few per cent of pelagic sediment. In contrast, Richardson *et al.*²² suggested that a similar component in the Walvis Ridge basalts was derived from enriched (metasomatized) material comparable to that proposed for the continental lithosphere.

A feature of the Dupal anomaly in the South Atlantic is that it appears to be restricted to an area between those of continental volcanism associated with continental break-up, Parana and Etendeka (Fig. 3). Moreover, both the HPT and the more primitive LTP rocks have the relatively high $^{87}\text{Sr}/^{86}\text{Sr}$ and $^{207}\text{Pb}/^{204}\text{Pb}$ ratios which define the Dupal anomaly in oceanic areas. LTP rocks, however, have minor and trace element features markedly different from OIB (for example, low Ti), which precludes any simple relation between the LTP rocks and oceanic magmatism (see also Fig. 1 inset). By contrast, the HPT rocks have many similarities with OIB.

For all three isotope systems, and for $1/\text{Sr}-^{87}\text{Sr}/^{86}\text{Sr}$, it is remarkable that the HPT Parana rocks plot at the end of the Walvis Ridge trend (Figs 1 and 2). In addition, the HPT lie at the high-Ba/Nb end of a striking positive correlation between Ba/Nb and $\Delta 7/4$ for South Atlantic basalts in the Dupal area²³. Thus, HPT rocks were either derived from source regions similar to the enriched component invoked for the Walvis Ridge, or they at least contain more of that component. This component has distinctive low $^{206}\text{Pb}/^{204}\text{Pb}$, similar to that inferred for portions of the sub-continental mantle from studies of lamproites and mantle xenoliths^{24,25}.

The Dupal signature has been identified in ocean islands such as Gough and Tristan da Cunha, on the Walvis Ridge², and in the HPT rocks of the Parana. It has been suggested that it was derived from either enriched lithosphere²², or recycled sediments²¹, but since both probably reflect subduction processes they may prove difficult to distinguish.

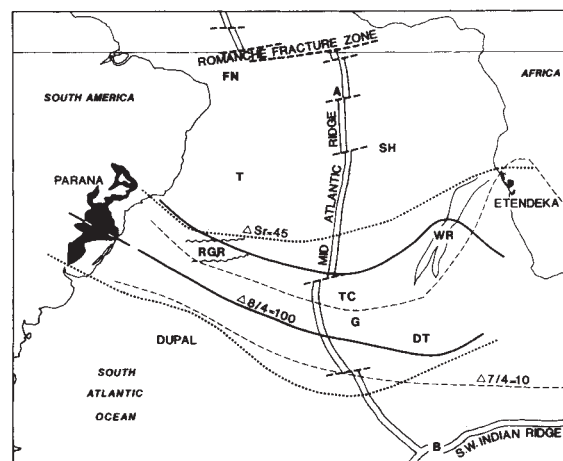


Fig. 3 Sketch map of Dupal anomaly isotope contours in the South Atlantic, after ref. 2. $\Delta 7/4$ and $\Delta 8/4$ are measures of the deviation of individual Pb isotope analyses from the NHRL, (Fig. 2). $\Delta 7/4 = [^{207}\text{Pb}/^{204}\text{Pb}_s - ^{207}\text{Pb}/^{204}\text{Pb}_{\text{NHRL}}]100$; $\Delta 8/4$ is calculated similarly, but $\Delta 8\text{r} = [^{87}\text{Sr}/^{86}\text{Sr}_s - 0.7]10^4$.

The presence of Dupal features in both oceanic and continental flood basalts, and their apparent restriction in the South Atlantic to areas between the Etendeka and the Parana, implies a close link between continental and oceanic magmatism. One interpretation is that the Dupal geochemistry is due to a deep-seated upwelling³ that both initiated continental break-up and was responsible for the Dupal signature. However, that requires that the trace element and isotope characteristics of the Karoo and Parana basalts are from deep-seated sources, rather than from within the continental lithosphere as presently believed^{4,10,11}, and it would conflict with suggestions that the mantle sources of both Karoo and Parana basalts are locally of similar age to the overlying crust^{12,12,16}. An alternative interpretation is that, at least in the South Atlantic, the Dupal is a relatively shallow-level phenomenon which evolved in the continental lithosphere. In this model, the Dupal signature occurs between the areas of continental volcanism because in those areas lithospheric mantle had been significantly heated and remobilized before continental break-up. Remobilization of such material detaches it from the lithospheric plates and hence enables it to contribute to oceanic volcanism as the continents move apart and the ocean basin develops.

We thank Peter van Calsteren, Andrew Gledhill and Roy Goodwin for help with the isotope analyses, and Peter van Calsteren and Nick Rogers for comments on the manuscript. Borming Jahn and Jon Patchett provided thoughtful reviews. Travel grants for M.M. were provided by FAPESP, the British Council and CNPq, and for this work the laboratories at the Open University and Oxford were supported by NERC.

Received 18 November 1985; accepted 22 April 1986.

1. Dupre, B. & Allègre, C. J. *Nature* **303**, 142-146 (1983).
2. Hart, S. R. *Nature* **309**, 753-757 (1984).
3. Cordani, U. G., Sartori, P. L. & Kawashita, K. *Anl. Acad. brasil. Cienc.* **52**, 811-818 (1980).
4. Erlank, A. J. *et al. Geol. Soc. S. Afr. spec. Publ.* **13**, 195-245 (1984).
5. Rabinowitz, P. D. & La Breque, J. J. *Geophys. Res.* **84**, 5973-6002 (1982).
6. Bellieni, G. *et al. J. Petrol.* **25**, 579-618 (1984).
7. Atalla, L. T. *et al. IAVCEI- IAGC Reykjavik* **52**, 62 (1982).
8. Mantovani, M. S. M. *et al. J. Petrol.* **26**, 187-209 (1985).
9. Bellieni, G. *et al. Neues Jahrbuch. Miner. Abh.* **150**, 273-306 (1984).
10. Cox, K. G. in *Continental Basalts and Mantle Xenoliths* (eds Hawkesworth, C. J. & Norry, M. J.) 139-157 (Shiva, Orpington 1983).
11. Duncan, A. R., Erlank, A. J., March, J. S. & Cox, K. G. *Geol. Soc. S. Afr. spec. Publ.* (1984).
12. Hawkesworth, C. J. *et al. Geol. Soc. S. Afr. spec. Publ.* **13**, 341-354 (1984).
13. Hawkesworth, C. J. *et al. in Continental Basalts and Mantle Xenoliths* (eds Hawkesworth, C. J. & Norry, M. J.) 111-138 (Shiva, Orpington 1983).
14. Patchett, P. J. *Nature* **283**, 559-561 (1980).
15. Huppert, H. E. & Sparks, R. S. J. *Earth planet. Sci. Lett.* **74**, 371-386 (1985).
16. Mantovani, M. S. M., Hawkesworth, C. J., Taylor, P. N. & Palacz, Z. *Abstr. Vol. IAVCEI—N.Z.* **181** (1986).
17. Leeman, W. P. *Lunar planet. Sci.* **XIII**, 431-432 (1982).

18. Cohen, R. S., O'Nions, R. K. & Dawson, J. B. *Earth planet. Sci. Lett.* **68**, 209–220 (1984).
19. Hofmann, A. W. & White, W. M. *Earth planet. Sci. Lett.* **57**, 421–436 (1982).
20. McKenzie, D. P. & O'Nions, R. K. *Nature* **301**, 229–231 (1983).
21. Weaver, B. L., Wood, D. A., Tarney, J. & Joron, J. L. *J. geol. Soc. Lond.* (in the press).
22. Richardson, S. H., Erlank, A. J., Duncan, A. R. & Reid, D. L. *Earth planet. Sci. Lett.* **59**, 317–342 (1982).
23. Hawkesworth, C. J., van Calsteren, P., Rogers, N. W. & Menzies, M. A. in *Mantle Metasomatism* (eds Menzies, M. A. & Hawkesworth, C. J.) (Academic, in the press).
24. Kramers, J. D. *Earth planet. Sci. Lett.* **42**, 58–70 (1979).
25. Le Roex, A. P. *et al.* *J. Petrol.* **24**, 267–318 (1984).
26. Erlank, A. J. *et al.* in *Mantle Metasomatism* (eds Menzies, M. A. & Hawkesworth, C. J.) (Academic, in the press).
27. Le Roex, A. P. *et al.* *J. Petrol.* **24**, 267–318 (1984).
28. O'Nions, R. K., Hamilton, P. J. & Evensen, N. M. *Earth planet. Sci. Lett.* **34**, 13–22 (1977).
29. Sun, S. S. *Phil. Trans. R. Soc. A* **297**, 409–445 (1980).

Acidity of Scottish rainfall influenced by climatic change

T. D. Davies*, P. M. Kelly†, P. Brimblecombe*,
G. Farmer† & R. J. Barthelmie†

* School of Environmental Sciences, and † Climatic Research Unit,
School of Environmental Sciences, University of East Anglia,
Norwich NR4 7TJ, UK

Although there are clear links between acidic deposition and synoptic meteorology¹, the effects of climatic change have been largely neglected. It is demonstrated here that long-term variations in the atmospheric circulation and associated changes in trajectory characteristics on timescales of years and longer can affect deposition levels within the United Kingdom, masking the effects of changing emissions. This study provides empirical support for the suggestion that climatic change may need to be considered in the assessment of emission control strategies².

The potential importance of climatic change in influencing precipitation acidity has been acknowledged^{3–6} but not assessed, although significant changes in climate and the atmospheric circulation have occurred in the European area over the past 100 yr (refs 7–9). As the frequency and character of air trajectories between regions varies, the relative importance of different source regions for deposition in a particular area will alter. Moreover, other physical and chemical processes in the atmosphere which influence deposition may also be affected. The spatial pattern of deposition over Europe as a whole could alter and climatic change may explain the apparent disagreement between recent trends in European pollutant emissions and precipitation composition^{10,11}.

Here, we consider the link between recent fluctuations in precipitation acidity (H^+ concentration) at Eskdalemuir in Scotland (Fig. 1) and variations in the frequency of various atmospheric circulation types over the United Kingdom extracted from the Lamb Catalogue of Daily Weather Types¹². This catalogue provides a succinct characterization of the synoptic circulation and extends back to 1861 enabling the study of long-term trends. On an annual basis, three Lamb Weather Types (LWTs) account for much of the variance in the circulation over the United Kingdom: Westerly (W); Anticyclonic (A); and Cyclonic (C)⁸. A-types produce little rainfall although they may be of ultimate importance for acidic deposition¹. C- and W-types together produce ~50% of the precipitation in the UK; no other type accounts for more than ~10%.

We have undertaken trajectory analyses for the C- and W-types to determine their significance for acidic deposition at Eskdalemuir. The synoptic situation characterized by the C-type¹² is such that, when there is frontal precipitation over Scotland, back-trajectories often originate over England or adjacent continental Europe (work in preparation). This is likely to lead to relatively acidic precipitation over Scotland^{13,14}. W-types, on the other hand, are more commonly associated with maritime trajectories and less-polluted rainfall in Scotland¹³. Variation in

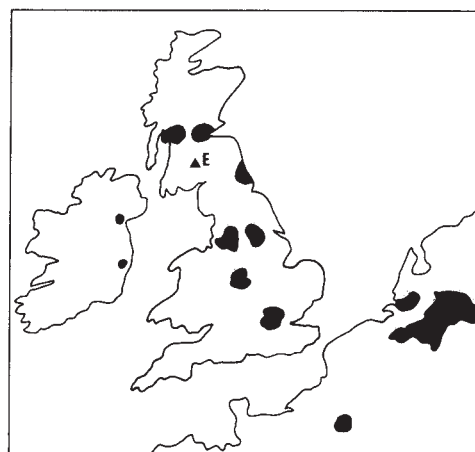


Fig. 1 Location of precipitation-collecting station used in this study: E, Eskdalemuir. The shaded areas represent major pollution sources. H^+ concentration data available for this station include: a, bulk-monthly, 1958–78 (Meteorological Office); b, daily, 1973–77 (Warren Spring Laboratory); c, daily, November 1977 to September 1982 (EMEP), same collector as b. Further details are available elsewhere^{14,20}.

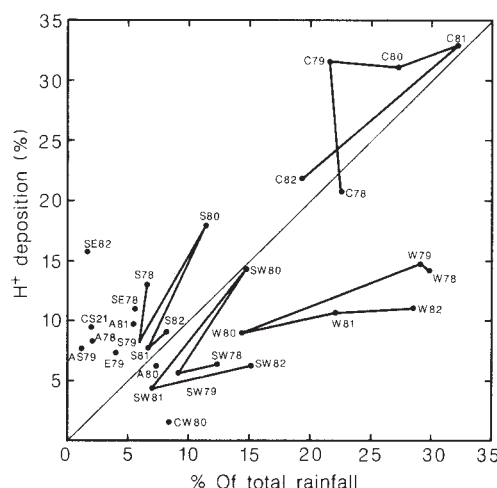


Fig. 2 Relationship between Lamb Weather Type, percentage of total H^+ deposition and precipitation amount at Eskdalemuir (annual values derived from EMEP daily data). Only LWTs contributing >7.5% of annual H^+ deposition and >7.5% of annual precipitation have been included. For clarity, the points for the major LWT contributors have been joined. The acid content of annual precipitation is dominated by C-type and W-type events. C-type events account for about 30% of the total H^+ deposition and around 25% of the total rainfall; the respective values for W-type events are ~12% and 25%. Although other LWTs make significant contributions to the concentration or dilution of precipitation acidity in certain years (S-type and SE-type events in particular), the acid content of annual precipitation is dominated by C- and W-type associated rainfall. C-type events make the major contribution of all LWTs to concentrating H^+ levels (a large average departure in the +y direction from the equality line combined with a substantial contribution to the total rainfall) whereas W-type rainfall makes the major contribution to diluting H^+ levels (greatest departure in the -y direction from the equality line along with a large contribution to the annual rainfall). C, Cyclonic; CS, cyclonic-southerly; CW, cyclonic-westerly; W, westerly; A, anticyclonic; S, southerly; SE, southeasterly; SW, southwesterly; E, easterly.

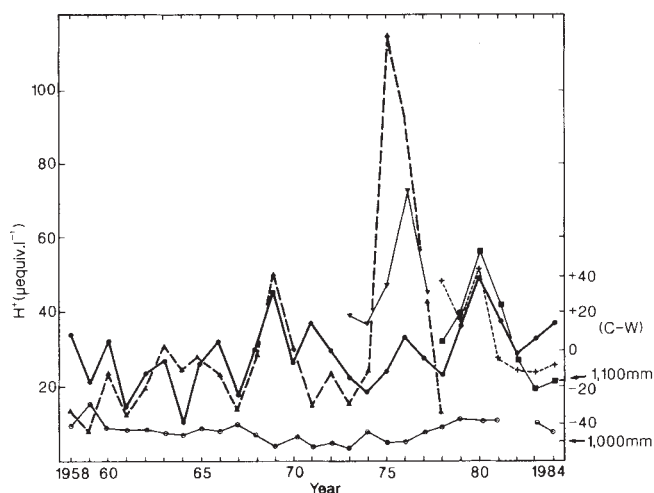


Fig. 3 Annual values of the C-W index, rainfall total for Eskdalemuir and volume-weighted H^+ concentration data for Eskdalemuir (for sources, see Fig. 1). The average H^+ concentration for seven collectors distributed through Scotland is also shown (data provided by J. N. Cape). The reliability of the monthly bulk collector data should be viewed with extreme caution²⁰. Unfortunately, they represent the only long-term records of precipitation acidity for the UK. Bottom curve, annual rainfall (mm); Δ and dashed line, Eskdalemuir monthly H^+ ; ∇ and light solid line, Eskdalemuir daily H^+ ; $\blacksquare$ EMEP daily H^+ ; $\bullet$ and heavy solid line, C-W; + and dotted line, Scotland regional H^+ .

wind speed between these types may also affect precipitation acidity¹³, as may other characteristics of the weather types. On the basis of our analysis of trajectories associated with the two LWTs, we conclude that these indicators of the general state of the local circulation can be used as a surrogate for case-related trajectory analysis; an expedient which allows us to explore the historical dimension of acidic deposition.

Analysis of the daily EMEP precipitation composition data¹⁴ for Eskdalemuir for the period January 1978 to October 1982 shows the prevalence of C-types in acidic episodes¹. These episodes, periods of high deposition on the day-to-day timescale, were first ranked according to order of magnitude. Then, the LWTs associated with the largest episodes each year (in total accounting for one-third of the total H^+ deposition in that year) were identified. The resulting frequency of C-types (~10 events per year) was four to five times that of the nearest contenders, the S-type (Southerly) and the W-type.

In determining the significance of the individual weather types for total H^+ deposition, precipitation amount is also important. Figure 2 illustrates the relationship between H^+ deposition, precipitation amount and LWT. C-type events account for between 20 and 25% of total H^+ deposition; W-types for 10–15%. C-type rainfall is high and relatively acid; W-type rainfall is similar in amount but lower in acidity. A simple index of the influence of atmospheric circulation variation on H^+ deposition can, therefore, be devised by subtracting the number of W-type days from the number of C-type days (the C-W index). This should be related to volume-weighted H^+ concentration in precipitation. More complex indices which included other LWTs and weighting to account for relative importance were devised but did not differ substantially in terms of information content from the C-W index.

Figure 3 shows the annual C-W index for the period 1958–84 during which rainfall H^+ concentration observations for Eskdalemuir are available. Despite problems with the reliability of the earlier acidity data, the overall relationship between these variables is clear, confirming that shifts in the atmospheric circulation do play a part in the determination of deposition

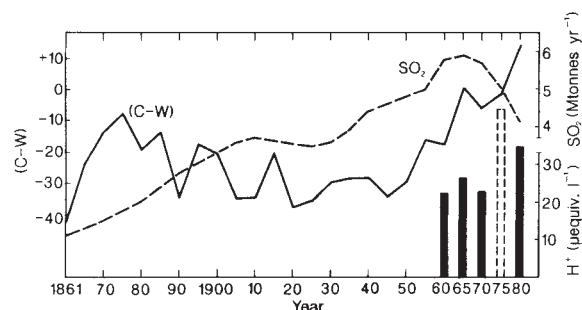


Fig. 4 Five-year means for: the C-W index (1861–64, 1865–69, and so on); SO_2 emissions estimated from published data²² (the data must be considered notional before 1950); and 5-yr volume-weighted H^+ concentrations at Eskdalemuir using the Meteorological Office bulk monthly data and the EMEP daily data. In the latter case, the pentade 1975–79 is represented differently to highlight the anomalous nature of the 1975–76 period; and the use of two different data sources during this interval.

levels in the Eskdalemuir region.

The relationship breaks down in certain years: for example, 1971 and, most notably, 1975 and 1976. The period of greatest acidity during the latter years, June 1975 to July 1976, was coincident with a severe drought over England and Wales^{15–17}. Frontal depressions were steered further north over the UK than usual and, although there was a drought in the South, Scotland received between 70 and 110% of its 'normal' rainfall. Synoptic patterns which are recognized to produce acidic episodes frequently affected Scotland. Ozone concentrations were also high over the UK during this period¹⁸ and this may have influenced sulphate production¹⁹ and led to greater rainfall acidity, further complicating the situation.

The C-type count for the 1975–76 period was low and, in this case, the simple C-W index was not a good measure of the atmospheric factors affecting production, transport and deposition. A more sophisticated index may be needed. Nevertheless, the correlation between the annual C-W index and the H^+ concentration data for the period 1958–84 is statistically significant. A correlation analysis yields a coefficient of 0.38, significant at the 5% level (the coefficient is 0.59 if the years of 1975 and 1976 are excluded). The relationship between these variables cannot be stationary in time because of changes in other important factors such as varying emissions. A more appropriate measure of association may, therefore, be a sign test on the year-to-year fluctuations. In this case, the test statistic is significant at the 0.1% level (data were included for all years during the period 1958–84).

Figure 4 shows the long-term history of the C-W index. Allowing for the anomaly in the mid-1970s, the 5-yr volume-weighted H^+ concentrations at Eskdalemuir show a closer match with this record than with that of UK emissions of SO_2 . (Sulphate has been the anion best correlated with precipitation acidity in Scotland²⁰.) Besides the strong relationship on the year-to-year timescale, there may also be a link on longer timescales. Unfortunately, the H^+ record is too short to test this point statistically.

This association between climatic change and precipitation acidity in the UK shows that the climate dimension must be taken into account in any consideration of the acid rain issue. Our index could be criticised on the grounds of oversimplicity and further work is needed to determine whether other indicators are more appropriate. Nevertheless, this analysis demonstrates that, as climate changes, deposition levels can vary even in the absence of alterations in emissions. As well as compounding the problem of assessing emission control strategies², this complicates the determination of the effects of past emissions. Climatic change may produce biological damage in its own right

or in conjunction with pollutants²¹, and it may also exert an indirect influence via its contribution to acidic deposition variability.

This work is funded by the UK Department of Environment and the Commission of the European Community. We thank Dr J. N. Cape (Institute of Terrestrial Ecology), Mr A. Martin (Central Electricity Generating Board), Dr J. G. Irwin (Warren Spring Laboratory) and Dr A. F. Tuck (Meteorological Office) for helpful advice and for permission to use their data. We also thank Dr F. B. Smith (Meteorological Office) and Dr B. E. A. Fisher (CEGB) for their valuable comments. The Norwegian Institute for Air Research was particularly helpful with the provision of data as were the Central Electricity Research Laboratories.

Received 15 January; accepted 2 May 1986.

- Smith, F. B. & Hunt, R. D. *Atmos. Environ.* **12**, 461–477 (1978).
- Streets, D. G., Lesht, B. M., Shannon, J. D. & Veselka, T. D. *Envir. Sci. Technol.* **19**, 887–893 (1985).
- Munn, R. E. & Rodhe, H. *Tellus* **23**, 1–13 (1971).
- Fisher, B. E. A. in *Sulphur in the Environment* (ed. Nriagu, J. O.) 243–295 (Wiley, New York, 1978).
- Ottar, B. *Atmos. Environ.* **12**, 445–454 (1978).
- Rodhe, H. & Granat, L. *Atmos. Environ.* **18**, 2627–2640 (1984).
- Lamb, H. H. *Climate: Present, Past and Future* Vol. 2 (Methuen, London, 1977).
- Jones, P. D. & Kelly, P. M. *J. Climat.* **2**, 147–157 (1982).
- Hess, P. & Brezowsky, H. *Ber. Deut. Hett.* **15**, No. 113 (1969).
- Cape, J. N. *et al. Atmos. Environ.* **18**, 1921–1932 (1984).
- Clarke, P. A., Fisher, B. E. A. & Scriven, R. A. *Atmos. Environ.* (in the press).
- Lamb, H. H. *Geophysical Memoir* No. 116 (HMSO, London, 1972).
- Fowler, D. & Cape, J. N. *Atmos. Environ.* **18**, 1859–1866 (1984).
- Schaug, J., Dovland, H. & Skjeltmoen, J. E. *ECE Co-operative Programme for Monitoring and Evaluation of the Long-range Transmission of Air Pollutants in Europe, Data Reports, October 1977–September 1982* (Chemical Co-ordinating Centre, Norwegian Institute for Air Research, Oslo, various years).
- Murray, R. *Met. Mag.* **106**, 129–145 (1977).
- Ratcliffe, R. A. S. *Met. Mag.* **106**, 145–154 (1977).
- Doomkamp, J. C., Gregory, K. J. & Burn, A. S. (eds) *Atlas of Drought in Britain 1975–76* (Institute of British Geographers, London, 1980).
- Simpson, D. *Report LR 536 (AP)M* (Warren Spring Laboratory, Stevenage, 1985).
- Martin, A. & Barber, F. R. *Atmos. Environ.* **19**, 1091–1102 (1985).
- UK Review Group on Acid Rain, *Acid Deposition in the UK* (Warren Spring Laboratory, Stevenage, 1983).
- Binns, W. O. & Redfern, D. B. *Forestry Commission Research and Development Pap.* 131 (Forestry Commission, Edinburgh, 1983).
- Irwin, J. G. *Envir. Hlth* **93**, 212–220 (1985).

Simultaneous outflow of fresh water and inflow of sea water in a coastal spring

Claude Drogue & Pascal Bidaux

Laboratoire d'Hydrogéologie, Université des Sciences et Techniques, Place E. Bataillon, 34060 Montpellier Cedex, France

The ability of sea water to flow into coastal or submarine cavities is surprising, in view of the fact that sea level is, for water, a level of minimum gravitational potential where it can perform no work. To achieve a lower potential requires energy, the provision of which is an example of a completely original aspect of the interaction between subterranean hydrodynamics and marine hydraulics. We have studied a spectacular example of the simultaneous outflow of fresh water and inflow of salt water at a karstic spring running into a salt-water lagoon on the French Mediterranean coast, and we suggest here a hypothesis to account for the surprising double-layered, two-directional flow observed.

The flow of sea water into cavities is known to occur in, among other places, the Bahamas, Florida and Greece, where limestone which is in contact with the sea displays considerable karst development and is very permeable. There appear to be various causes—proved or supposed—for the inflow of sea water^{1–5}; the examples described in the literature are continuous or periodic but consist of either inflow of sea water or outflow of fresh water. The two types of flow never occur simultaneously.

The work described here reveals the possibility of a more complex coastal phenomenon in which two flows—outflow and

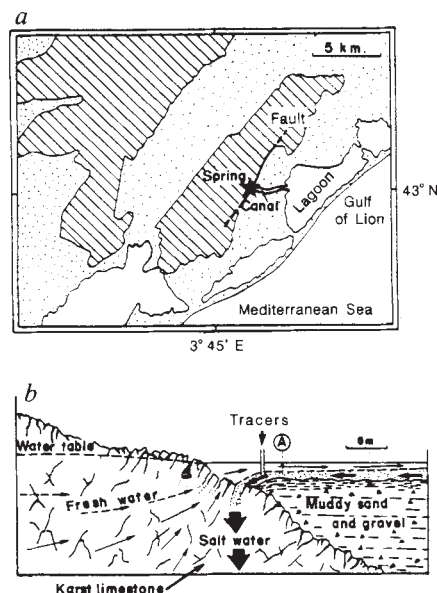


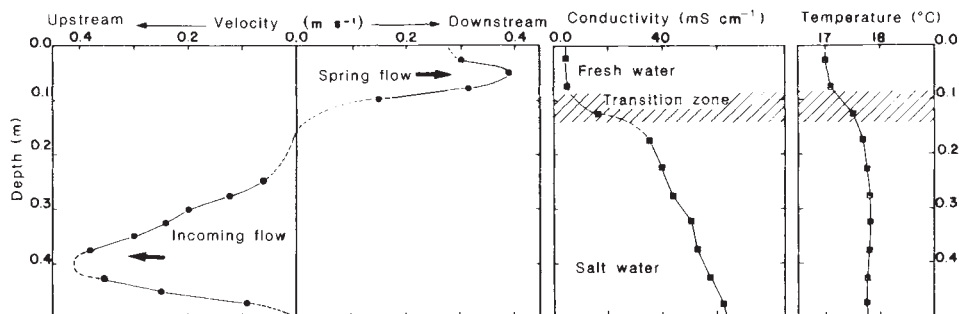
Fig. 1 *a*, Map showing the location of the Source de la Roubine. The carbonate massif has an area of 55 km² and consists of Mesozoic limestone and dolomite. It is only part of the aquifer feeding the spring; other massifs located further north also contribute to the supply. The deposits of Quaternary and Cenozoic clay, loam and sand in the plain are several hundred metres thick and cover Mesozoic limestone which extends beneath the sea. The level of the Mediterranean has undergone a number of fluctuations since the end of the Neogene, and during some periods was lower than it is today^{10–13}. These regressions caused deep penetration by rainfall infiltrating into the carbonate massifs which had emerged, accompanied by intense internal corrosion and the formation of cavities by dissolution (karst topography). Thus, the aquiferous structure which formed in the depths of the carbonate mass is mainly below sea level today, and the Source de la Roubine is a spring of very recent formation. Movement of water starts in the form of a rising current in the deep palaeokarst. *b*, Cross-section of the spring showing the double flow (vertical exaggeration 5:1). Tracers were injected into the stream of salt water through a pipe, as shown in the drawing. Measurements of velocity, conductivity and temperature were made along the vertical marked 'A'.

inflow—are concomitant and superposed. The study was carried out at a spring on the Mediterranean coast (Source de la Roubine, Fig. 1*a*), which is at the edge of an extremely karstic carbonate massif and quaternary clay-marl deposits several hundred metres thick in the coastal plain. The spring and flow run along a 2.5-km channel into a lagoon (sea water). The discharge can attain a value of 6 m³ s⁻¹, and is 0.1–0.2 m³ s⁻¹ at normal low-water functioning of the spring, that is, when the spring is flowing. Inflow of salt water occurs only during such times of low discharge⁶.

In the absence of rain, rises of several tens of centimetres are observed in the water level near the spring and last from a few hours to 4–5 days. While the water level changes in this way, a current of salt water runs along the bottom of the channel below the fresh water and in the opposite direction, enters one or two of the springs and disappears into the aquifer. At the same time, water continues to run out of the other springs and flows towards the lagoon on top of the salt water (Fig. 1*b*).

The interface between the two flows can be clearly seen in vertical profiles of electrical conductivity and temperature. (Fig. 2). The conductivity of the water from the spring is 4.2–4.3 mS cm⁻¹ (at 25 °C), whereas that of the lagoon water varies from 48 to 56 mS cm⁻¹ according to the season. Although for reasons of convenience the spring water is referred to as 'fresh', the mean electrical conductivity of spring water in limestone soil far from the coast is 0.6 mS cm⁻¹ with a Cl⁻ ion content of

Fig. 2 Observations of velocity, conductivity and temperature, showing the superposition of the two streams. Maximum velocities of the two currents running in opposite directions are almost the same. The discharge of the fresh-water stream is $0.18 \text{ m}^3 \text{ s}^{-1}$; that of the salt-water stream beneath is $0.20 \text{ m}^3 \text{ s}^{-1}$. The superposition of fresh water is also shown by the observations of conductivity and temperature. The lagoon water is warmer than the spring water.



$15\text{--}35 \text{ mg l}^{-1}$ (the water in the lagoon has a Cl^- content of $18,000 \text{ mg l}^{-1}$).

The temperature of the spring is relatively stable, at $17.3 \pm 0.5^\circ\text{C}$. In contrast, the lagoon water is subject to wide temperature variations: from $5\text{--}10^\circ\text{C}$ in winter to $25\text{--}28^\circ\text{C}$ in summer. The monitoring of inflow with time revealed some remarkable features. The beginning of the double-flow phenomenon has a sudden effect on conductivity values: this is proof of the poor miscibility of fresh and salt water. The temperature, however, varies progressively even before the appearance of salt water. This is caused by the fact that the advance of the salt water in the bottom of the channel pushes the fresh spring water already affected by the outside temperature back towards the spring.

With regard to the hydraulic mechanism, it would appear that the variations in the water level at the spring are caused by variations in the level of the lagoon on the west bank. The measurements (Fig. 3) show that they are not caused by tide;

however, it was observed that the variations are associated with certain wind directions. The spring is on the west bank; easterly winds cause the water level to rise on this side and fall on the east bank. The rapid rise in level on the western side thus sets up a hydraulic gradient in the salt water, directed towards the spring. Progress of salt water towards the spring is encouraged by the low hydraulic gradient of the fresh water in the channel ($\sim 0.01\%$) and by the near-zero slope of the bottom of the channel.

The superposition of the two flows and the low apparent miscibility of the two streams are caused by the differences in density (spring water: $0.9994 \text{ kg dm}^{-3}$ at 25°C ; sea water (in the lagoon): $1.0258 \text{ kg dm}^{-3}$).

Finally, it might be thought that the salt water which runs into the aquifer reappears in spring flow after dilution and could be the cause of the salinity of the spring. Tracers were injected into the saline inflow in order to test this hypothesis, and no trace of the dye was observed in fresh water to the spring during an observation period which lasted several months. It would thus appear that the two flows are probably separate in the aquifer itself.

It is likely that part of the salt water mixes with water from the aquifer, but the origin of most of the salinity of the spring water is undoubtedly contamination of deep ground water by the penetration of sea water into the karst⁷⁻⁹ (salt-water encroachment; Fig. 4). When lagoon water has flowed in, it probably runs by gravity to the interface of the salt-water encroachment. It may then be taken by groundwater flow to springs under the lagoon or under the sea.

In conclusion, we note that geological structures which could

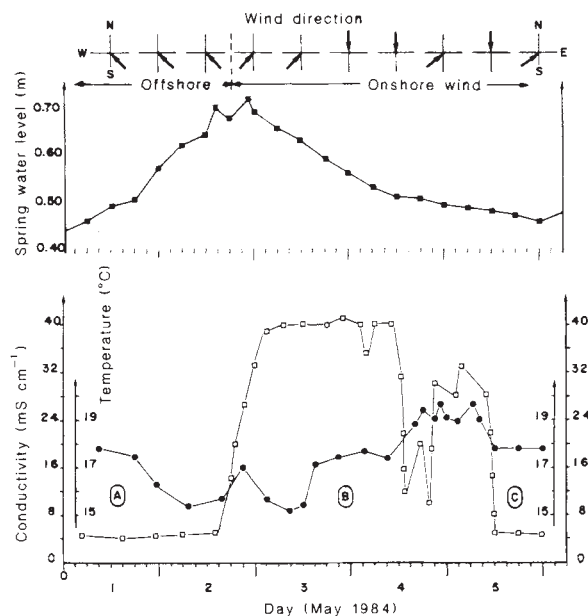


Fig. 3 Example of the time-evolution of conductivity ($\square$) and temperature ($\bullet$) in the spring into which the salt water flows, compared with water levels in the spring ($\blacksquare$) and with wind direction (arrows). A, No inflow of salt water, but start of return flow is indicated by a fall in temperature. This is fresh water in the channel affected by the outside temperature. B, Inflow has started suddenly and is revealed by stable conductivity; however, temperatures are irregular. This is the effect of the thermal heterogeneity of the water in the lagoon. The lowest temperature of the water flowing in is 14.9°C ; the highest is 19.8°C (the temperature of the spring water is 17.8°C). C, Inflow ceases suddenly. The spring flows normally once more. Inflow occurs when there is a rise in water level of 0.35 m caused by easterly winds.

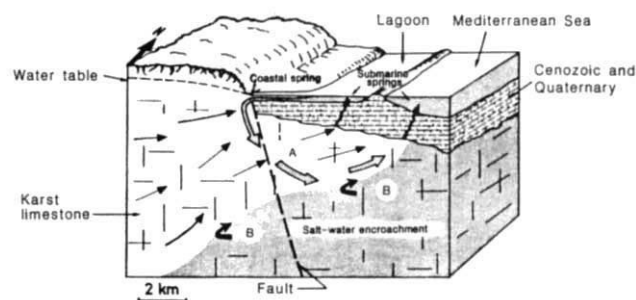


Fig. 4 Diagrammatic representation of the karstic aquifer and the organization of flows. Fine arrows, fresh water; bold and shaded arrows, salt water. The flow of salt water that runs into the spring must descend by gravity through the aquifer (A) to the deep salt-water encroachment (B). This flow is then carried to submarine springs. There is thus double contamination of underground water by sea water: from below in the aquifer and from above. Note that the large-scale flow of water into the spring carries large quantities of organic matter into the underground zone; this could form the basis for a totally unknown biotope. (For reasons of simplicity, the arrows in A do not show any dispersion that may take place in the aquifer.)

enable this phenomenon to occur are very common on the limestone coasts around the Mediterranean. Equivalent structures also exist on ocean coasts: we have observed such features in Miocene limestone in south-east Java (Gunung Sowu) and in southern Portugal (Algarve).

Received 27 January; accepted 28 April 1986.

1. Stringfield, V. T. & Legrand, H. E. *J. Hydrol.* **14**, 139–157 (1974).
2. Stringfield, V. T. & Legrand, H. E. *J. Hydrol.* **9**, 387–404 (1969).
3. Maurin, V. & Zoelt, J. *Int. Ass. Sci. Hydrol. Bull.* **74**, 423–438 (1967).
4. Fuller, M. L. *Geol. Soc. Am. Bull.* **18**, 221–232 (1906).
5. Geze, B. *La Géographie* **69**, 193–208 (1938).
6. Derozier, Ph. thesis, Univ. Montpellier (1984).
7. Kohout, F. A. *J. geophys. Res.* **65**, 2133–2141 (1960).
8. Henri, H. R. *Int. Ass. Sci. Hydrol. Bull.* **52**, 478–487 (1960).
9. Cooper, H. H., Kohout, F. A., Henri, H. R. & Glover, R. E. *U.S. Geol. Surv. Wat. Supply Pap.* No. 1613–C (1964).
10. Hsu, K. J. *et al. Init. Rep. DSDP* **42**, 1053–1078 (1978).
11. Steckler, M. S. & Watts, A. B. *Nature* **287**, 425–429 (1980).
12. Aloisi, J. C., Monaco, A. & Thommeret, J. *Rev. Géogr. phys. Geol. dynam.* **2**, 13–22 (1975).
13. Monaco, A. & Thommeret, J. *C. r. heb. Séanc. Sci., Paris.* **D274**, 2280–2283 (1972).

Sign language aphasia during left-hemisphere Amytal injection

Antonio Damasio*, Ursula Bellugi†‡, Hanna Damasio*, Howard Poizner† & John Van Gilder*

* Department of Neurobiology, University of Iowa College of Medicine, Iowa City, Iowa 52242, USA

† Laboratory for Language and Cognitive Studies, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA

Although it has been established that the left cerebral hemisphere subserves spoken language, the nature of brain organization for sign language remains relatively unexplored. The issue is especially important because sign language displays the complex linguistic structure of spoken languages, but conveys it through manipulation of visuo-spatial relations¹, thereby exhibiting properties for which the hemispheres of hearing individuals show opposing specializations. We had the unique opportunity to study a hearing signer proficient in American sign language (ASL), during the left intracarotid injection of a barbiturate (the Wada test), and before and after a right temporal lobectomy. The subject was a strong right-hander. Neuropsychological and anatomical asymmetries suggested left cerebral dominance for auditory-based language. Emission tomography revealed lateralized activity of left Broca's and Wernicke's regions for spoken language. The Wada test, during which all left language areas were rendered inoperative, caused a marked aphasia in both English and ASL. After partial ablation of the right temporal lobe, the abilities to sign and understand signing were unchanged. These data add further support to the notion that anatomical structures of the left cerebral hemisphere subserve language in a visuo-spatial as well as an auditory mode.

Recent research has shown that ASL has developed as a fully autonomous language. It has complex organizational properties shared by the spoken languages of the world, but it also has grammatical devices not derived from spoken language; thus, ASL is a separate language, not a manual rendering of English. At all structural levels, the surface forms of a signed language are deeply influenced by the modality in which it develops. This is displayed most distinctively in the pervasive use of spatial contrasts and spatial manipulations to express syntactic function. At the syntactic level, for example, nouns are associated with arbitrary points in signing space, pronoun signs are directed toward those points for anaphoric reference, and verb signs move between them in specifying grammatical relations². Thus a grammatical function served in many spoken languages by case marking or by linear ordering of words is fulfilled in ASL by essentially spatial mechanisms. Since sign languages utilize

visuo-spatial channels, several investigators have assumed that structures in the right cerebral hemisphere are likely to have a major role in the reception and production of these languages. Indeed, the issue is quite controversial and many studies do suggest a special role for the right hemisphere in mediating signs³.

Our study was performed in a 27-yr-old hearing woman, a native English speaker who learned ASL at an age of 18 yr. She is a college graduate with a MS degree in rehabilitation of the deaf, and works at a community agency as interpreter and counsellor for deaf people. She uses ASL daily in her work and is a skilled signer. She had her first seizure, of the partial complex type, at age 13 yr, but had never had generalized tonic-clonic seizures. The seizures did not hamper her intellectual or emotional development, but in recent years the episodes had become more frequent and refractory to drug therapy. Because they were interfering with her professional activity she sought surgical treatment. She is a right-hander (in Geschwind's version of Oldfield's laterality index, she scored +90), and so are both her parents. Comprehensive neuropsychological evaluation failed to reveal any abnormality of language, visuo-spatial or other higher cognitive abilities. Her full-scale intelligence quotient (Wechsler Adult Intelligence Scale-Revised) was 105, with no discrepancy between the verbal and performance scales. Her dichotic listening performance for words showed a right-ear advantage. Computer tomographic scan and cerebral angiography revealed brain asymmetries suggestive of left language dominance (left > right occipital width on computer tomography; left sylvian angle = 105°; right sylvian angle = 130°, in angiography). Recording from depth electrodes showed a right anterior temporal focus. There was no mirror focus.

Single photon emission tomography (SPET) using xenon-133 as radiotracer revealed a normal resting pattern⁴. A second SPET procedure was performed during a language activation task, a verbal rhyme detection paradigm modified from Knopman⁵. The subject compared words presented through earphones with a memorized target word and was asked to respond with left foot flexion to those words that rhymed with the memorized target. SPET revealed increased signal in the left Broca and Wernicke regions (see Fig. 1), indicating that processing of English was lateralized to the left.

Injection of a barbiturate (sodium amobarbital; Amytal) was made into the left carotid system (due to a persistent right trigeminal artery, no injection was possible in the right side). The procedure was videotaped and both English and ASL examiners were in attendance. The patient developed a right hemiplegia and aphasia in both English and ASL. During the recovery period, correct English responses appeared at 4 min, preceding those in ASL by 1½ min, a substantial time interval in this procedure. The patient's signing was markedly impaired, including sign paraphasias, perseverations, neologisms and grammatical errors. Interestingly, since she was hearing and could sign and speak at the same time, we could compare her responses in two languages simultaneously. This revealed a frequent mismatch between word and sign, with English mostly correct but ASL often incorrect in both meaning and form (see Fig. 2).

The extent of the surgical resection was documented by post-operative advanced magnetic resonance imaging (see Fig. 3). It encompassed the right hippocampus, parahippocampal gyrus, amygdala and both polar and anterolateral right temporal neocortex. The primary and secondary auditory association cortices contained in the middle and posterior first temporal gyrus were spared, as was area 37.

Neuropsychological re-evaluation at 3 and 12 months comprised all the tests administered pre-operatively, to assess attention, perception, language, memory and reasoning. The scores were virtually unchanged, including those of the Boston Diagnostic Aphasia Examination⁶, dichotic listening, recognition and recall of previously learned visual stimuli, geographical orienta-

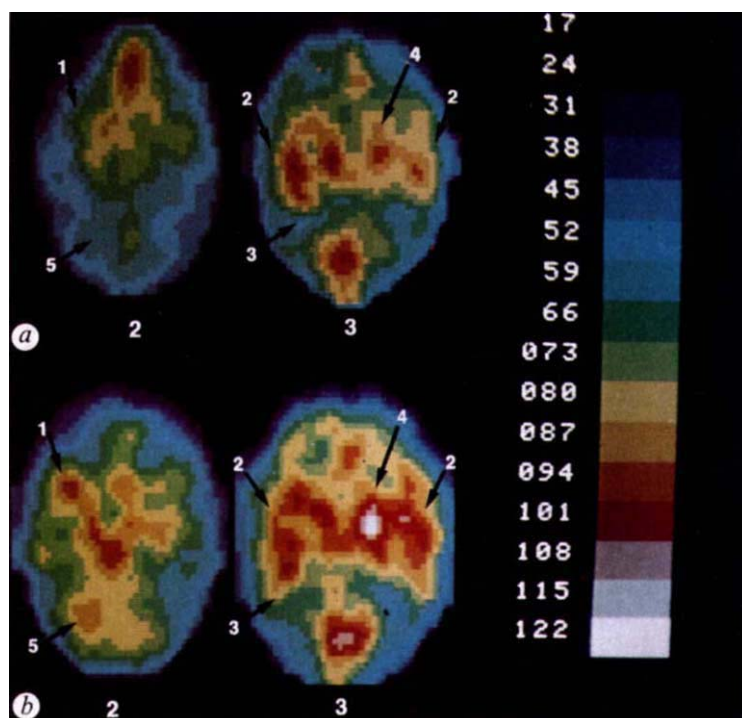


Fig. 1 Cuts 2 and 3 of a six-slice Tomographic SPET study at rest (*a*) and during a verbal rhyme detection task (*b*). During activation there is an increase of radiosignal in Broca's region, that is, in the left frontal fields 44 and 45 indicated by arrow 1, and a bilateral increase of signal in auditory cortices indicated by arrow 2. In the left hemisphere the signal increase extends into Wernicke's area, that is, the posterior sector of area 22 (arrow 3). During activation there is also an increase in the right basal ganglia (arrow 4) and left cerebellar hemisphere (arrow 5), related to the motor activity of the left foot required to signal appropriate rhyme detection.

tion, immediate visual memory (Benton's visual retention test), and copy and 15-min recall of a visual pattern (Rey-Osterrieth figure). The learning of new visuo-spatial and visual object memories was not assessed. There were no detectable neurological abnormalities. Most importantly, however, analysis of videotaped spontaneous signing and structured signing interviews confirmed the patient's own impression that her ability to sign and understand signing had not been compromised at all. Intensive analysis of her videotaped signing showed, for example, that her noun to pronoun ratio was in the normal range, her use of ASL morphological structures was unchanged, and there were no errors in her use of the spatially organized syntax of ASL before or after the right temporal lobectomy.

The impairment of ASL during the Wada test is associated with pharmacologically induced dysfunction in the territory of

the left middle and anterior cerebral arteries⁷. The structures clearly affected during the test included the left Broca and Wernicke areas, the left dorsolateral parietal lobe (especially the inferior parietal lobule), the middle and lower rolandic cortices, the left basal ganglia and possibly territories of both the lateral temporal lobe and the anterolateral frontal lobe⁸. The cerebral territories supplied by the left posterior cerebral artery (namely, the mesial and part of the lateral occipital lobe, and the mesial occipito-temporal and occipito-parietal regions) were probably unaffected. This indicates that a fully operational right hemisphere and partially operational, interconnected, left visual cortices could not prevent the impairment in sign language. Thus ASL may be especially dependent on the normal function of left hemisphere cortices, as is the case with auditory-based languages.

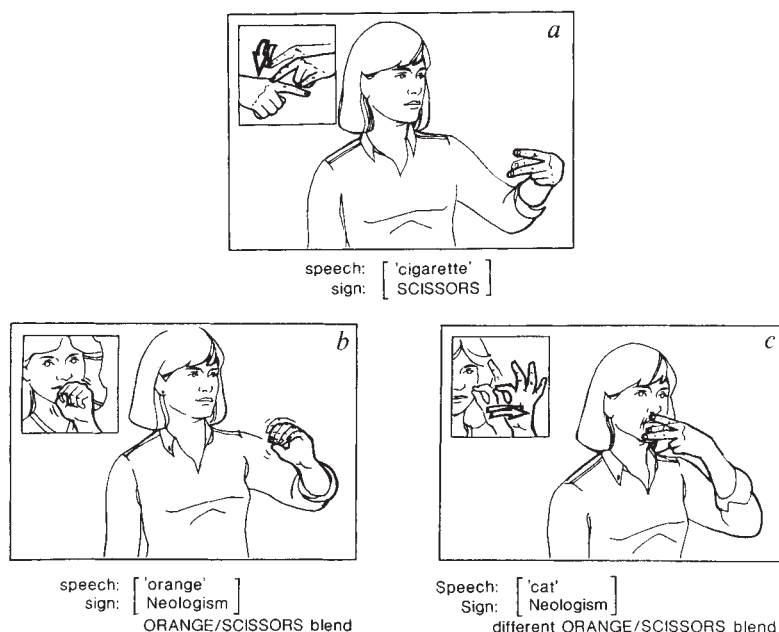


Fig. 2 Three errors produced by the patient during left Wada injection. In identifying an object presented, the patient produced an English word and an ASL sign simultaneously, and the two differed. In *a*, the patient correctly said 'cigarette' but simultaneously signed 'scissors' to a presented cigarette. In *b*, the patient correctly said 'orange' to a presented orange, but signed a form that was like a blend of two signs in ASL—the handshape of 'orange' and the place and movement of 'scissors'. In *c*, the patient correctly said 'cat' to a picture of a cat, but signed another blend, this time with the handshape of 'scissors', the place of articulation of 'orange' and the movement of 'scissors'. The combinations of parameters from different signs produced nonsense forms that were well-formed in ASL but were meaningless. It is of interest that during recovery from the left Wada injection, the patient frequently responded in speech and sign simultaneously—a unique possibility for languages in different modalities—with a mismatch between the two languages and the sign more often in error. Inserts in the upper left-hand corner of the illustrations indicate the correct ASL signs.

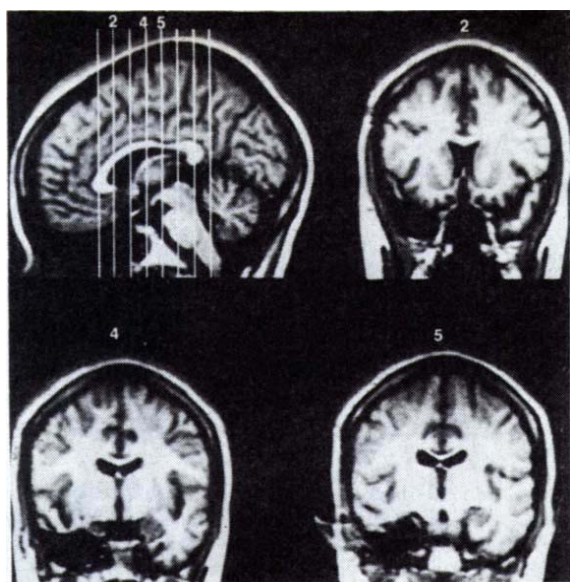


Fig. 3 Magnetic resonance images obtained after surgery with an inversion recovery technique; magnetic field, 0.5 T. Resolution was 2 mm in the plane of cut. The top left-hand image is a midsagittal cut showing the mesial aspect of one hemisphere. The vertical lines represent the level and incidence of the coronal cuts. Cuts 2, 4 and 5 are reproduced depicting the area of right temporal lobe ablation. Hippocampus, parahippocampal gyrus, fourth, third and second temporal gyri are missing. The first (superior) temporal gyrus is intact.

A major point of interest concerns the rate of recovery of the languages. During the final period of dysfunction, after the use of English had returned and the paralysis subsided, the patient still made errors in ASL. Considering that the recovery from barbiturate action must have taken place gradually and have occurred last in the outer border of the middle cerebral artery territory, that is, in a C-shaped watershed region containing visual and somatosensory cortices that we believe are especially crucial for ASL processing, it is possible that the ASL delay reflected the continued dysfunction of those cortices at a time when the core territories of the middle cerebral artery (which contain the classic speech cortices) had already returned to normal. An alternative explanation for this lag would need to invoke the fact that ASL was a second language and propose that its neural representation might thus be more vulnerable, a rather unsatisfactory hypothesis. In fact, one might have expected a second language, and especially a visuo-spatial language, to have been primarily represented in the non-English dominant hemisphere, in which case the left Wada test would not have interfered with ASL operations at all.

The lack of post-operative signing defects suggests that right anterior temporal structures medially and laterally do not have a significant role in signing or in comprehension of previously learned signs. It is possible that these areas are pertinent to the learning of ASL but certainly not to the production of ASL once learned, that is, the right hippocampal/parahippocampal complex and the polar and anterolateral temporal cortices were involved in neither the generation nor the detection of the complex visuo-spatial and motor patterns of ASL that had been learned pre-operatively. It might be argued that this patient's seizure would have so disorganized the structure of the right hemisphere that any ability normally dependent on it would have been transferred to the healthy left hemisphere. This possibility is unlikely considering: (1) the mild and circumscribed neuropathological changes encountered in the block of ablated tissue; (2) the lack of a mirror focus in the opposite hemisphere; (3) the lack of any pre-operative symptom or sign suggestive of right hemisphere dysfunction; (4) the late onset of the seizures;

and (5) the fact that they never interfered with the acquisition of skills or the development of a normal personality.

Our findings provide the rare opportunity of relating the processing of ASL to a circumscribed set of brain structures in the left hemisphere, in a subject who has unequivocal left-sided functional dominance for English, and who has anatomical brain asymmetries also favouring the left hemisphere. Furthermore, the findings suggest that the right hippocampal system is not involved in active processing of ASL. Our results should be seen in the perspective of other converging evidence on the respective roles of left and right hemisphere in ASL⁹⁻¹¹. Studies of unilateral lesions in deaf signers show that lesions to the left hemisphere cause sign aphasia, affecting different structural levels of the language, while lesions to the right hemisphere produce marked spatial deficits but leave expressive language intact¹². Thus, the underlying specialization of the left hemisphere for language does not seem to rest on speech or sound, nor on the form of the signal, but rather on the linguistic function it subserves.

We thank Dr Paul Eslinger for performing the neuropsychological assessment, Drs Karim Rezai and Val Dunn for their collaboration in the SPET and magnetic resonance imaging studies, Mr Ozzie Diaz-Duque and Ms Charissa Lansing for their assistance with sign language procedures during the Wada test and David Corina, Maureen O'Grady, Lucinda O'Grady and Diane Lillo-Martin for their help with the sign language analysis. This research was supported in part by NINCDS grant PO NS 19632-02 (to A.D. and H.D.), NSF grant BNS 83-09680 and NIH grants HD13249, NS15175, NS19096 (to U.B. and H.P.).

Received 15 October 1985; accepted 10 April 1986.

1. Klima, E. S. & Bellugi, U. *The Signs of Language* (Harvard University Press, 1979).
2. Bellugi, U. *Signed and Spoken Language: Biological Constraints on Linguistic Form*, 115-140 (Verlag Chemie, Weinheim, 1980).
3. Poizner, H. & Battison, R. in *Recent Perspectives on American Sign Language* (eds Lane, H. & Grosjean, F.) 79-101 (Erlbaum Associates, Hillsdale, 1980).
4. Rezai, K. *et al. Neurology* **34**(3), 204 (1984).
5. Knopman, D. S., Rubens, A. B., Klassen, A. C. & Meyer, M. W. *Arch. Neurol.* **39**, 487-493 (1982).
6. Goodglass, H. & Kaplan, E. *The Assessment of Aphasia and Related Disorders* (Lea & Febiger, Philadelphia, 1983).
7. Wada, J. & Rasmussen, T. *J. Neurosurg.* **17**, 266-282 (1960).
8. Damasio, H. *Arch. Neurol.* **40**, 138-142 (1983).
9. Hoeman, W., Criswell, E., Wada, A. & Ross, D. *Neurology* **32**, 1020-1023 (1982).
10. Bellugi, U., Poizner, H. & Klima, E. A. *Hum. Neurobiol.* **2**, 155-170 (1983).
11. Poizner, H., Kaplan, E., Bellugi, U. & Padden, C. *Brain Cognition* **3**, 281-306 (1984).
12. Bellugi, U., Poizner, H. & Klima, E. S. *What the Hands Reveal About the Brain* (Bradford Books, Cambridge, Massachusetts, in the press).

Electrokinetic shape changes of cochlear outer hair cells

Bechara Kachar*, William E. Brownell†, Richard Altschuler‡ & Jörgen Fex‡

Laboratories of * Neurobiology and ‡ Neuro-Otolaryngology, NINCDS, National Institutes of Health, Bethesda, Maryland 20892, USA

† Departments of Neuroscience and Surgery (ENT), Box J-244, University of Florida, Gainesville, Florida 32610, USA

Rapid mechanical changes have been associated with electrical activity in a variety of non-muscle excitable cells¹⁻⁵. Recently, mechanical changes have been reported in cochlear hair cells⁶⁻⁸. Here we describe electrically evoked mechanical changes in isolated cochlear outer hair cells (OHCs) with characteristics which suggest that direct electrokinetic phenomena are implicated in the response. OHCs make up one of two mechanosensitive hair cell populations in the mammalian cochlea; their role may be to modulate the micromechanical properties of the hearing organ through mechanical feedback mechanisms⁶⁻¹⁰. In the experiments

described here, we applied sinusoidally modulated electrical potentials across isolated OHCs; this produced oscillatory elongation and shortening of the cells and oscillatory displacements of intracellular organelles. The movements were a function of the direction and strength of the electrical field, were inversely related to the ionic concentration of the medium, and occurred in the presence of metabolic uncouplers. The cylindrical shape of the OHCs and the presence of a system of membranes within the cytoplasm—laminated cisternae¹¹—may provide the anatomical substrate for electrokinetic phenomena such as electro-osmosis^{12,13}.

Isolated OHCs were prepared from the organ of Corti of guinea pig by non-enzymatic mechanical dissociation^{7,8}. A 100- μ l volume of culture medium containing isolated cells was placed on a microscope slide and viewed with video-enhanced microscopy (Fig. 1). Isolated OHCs attach to the microscope slide at either the synaptic or the stereociliar end. Sinusoidal potential gradients at 2–30 Hz and <1 mA (0.8–1.5 V peak-to-peak) were passed across OHCs by means of a pair of silver-wire electrodes placed 200–300 μ m apart (Fig. 1). About 100 isolated OHCs from 34 cochleae were analysed from experiments recorded on videotape.

Oscillatory movement of the cell's free end, corresponding to elongation and shortening of the cell in response to the electrical stimulation, was visualized in over 80% of the OHCs. Oscillatory movement of intracellular organelles located around the cell axis was observed in ~15% of the OHCs. Maximal responses of up to 0.3–0.4 μ m amplitude about a resting position (Figs 2, 3) were obtained when the cells were longitudinally aligned with the direction of the electrical field. We could not distinguish oscillatory movements from brownian movement for amplitudes smaller than 50 nm. Most OHCs showed elongation with positive potential gradients and shortening with negative potential gradients, when the ground electrode was placed near the end of the cell that was attached to the substrate (Figs 2a–d, 3a,b,d). Less than 20% of the OHCs showed the opposite response polarity. Organelle movement, when detectable in cells that also responded with length changes, was in-phase with the movement of the free end of the cell (Fig. 3d,e). Organelle movement was also observed in some cases when both the stereociliar and synaptic ends were attached to the slide. Oscillatory movements of organelles or changes in cell length were not observed in red blood cells, Deiters' or Hensen's cells which co-isolate with the OHCs.

These electrically evoked responses persisted for several hours after dissociation. The responses were unchanged when the cells were incubated for 1 h in medium containing dinitrophenol (2 mM) or iodoacetic acid (2 mM), both of which are inhibitors of ATP production and commonly used to test for energy requirements^{14,15}. The responses were enhanced by lowering the ionic concentration by 20–25% (this was achieved by isotonic dilution of the medium with a sucrose solution). Some cells showed little or no movement in the standard medium but exhibited robust movement after ionic dilution; this effect was reversed by replacing the diluted medium with the original medium (Fig. 3d).

The symmetry of the sinusoidal displacements, their apparent independence of ATP involvement, and the enhanced responses after ionic dilution are difficult to reconcile with conventional contractile mechanisms. However, electrophoresis^{13,16} and electro-osmosis^{12,13} might provide the motive force underlying both the length changes and the oscillatory organelle movements. Such electrokinetic phenomena are a function of applied voltage and are enhanced in medium of low ionic concentration^{12,17}. Electro-osmosis is a field-induced movement of fluid adjacent to a charged surface as a consequence of the electrophoretic migration of counterions^{12,13,17–19}. When electro-osmosis occurs in a tube with closed ends, a pressure difference develops between the tube ends¹⁹. If the tube wall is elastic it can deform (elongate or shorten), or if it is rigid the flow of liquid along

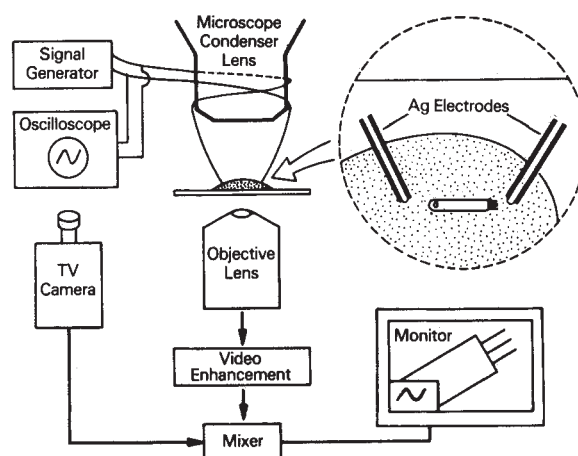


Fig. 1 Schematic diagram of the video-enhanced light microscopy³¹ system used in the present experiments. An inverted Zeiss Axiomat was operated in the differential interference contrast mode and equipped with a $\times 100$, 1.3-NA Planapo objective, a 0.63-NA condenser fully illuminated with a 100-W mercury lamp and a Newvicon Dage-68 (MTI) video camera for contrast enhancement. The magnification of the image on the TV monitor was $\times 20,000$. Experiments were carried out in droplets of medium (Leibovitz L-15 with 1% bovine serum albumin and 10 mM HEPES). A droplet containing cells was placed on a microscope slide which had been pretreated with polylysine. Isolated OHCs in most cases adhere to the slide surface at either the basal (synaptic) or apical (stereociliar) region. Stimulation was achieved with a pair of Teflon-insulated no. 30 silver-wire electrodes placed 200–300 μ m apart and mounted on the microscope condenser lens. The electrode pair was lowered into the droplet and oriented in relation to individual OHCs by rotating the microscope stage. The voltage waveform was displayed on an oscilloscope screen and the video image mixed with that of the video-enhanced image of the OHC. The response of the cells to electrical stimulation were analysed by frame-by-frame playback of videotape-recorded experiments. Time resolution was limited to 17 ms, corresponding to the duration of a single non-interlaced video frame. Photographs and measurements were taken from a TV monitor.

the wall is counterbalanced by a flow in the opposite direction at the centre of the tube¹⁸.

OHCs display structural features that are favourable for electrokinetic effects. OHCs are elongated cylinders, 4–9 μ m wide and 15–70 μ m long, and they contain parallel membranes that form the laminated cisternae along the lateral plasma membrane¹¹ (Fig. 2f). The function of the cisternae is unknown. The closely apposed membranes may be an optimal substrate for an electro-osmotic effect¹². Cytoplasmic fluid movement in relation to the plasma and cisternal membranes would produce pressure variations and changes in cell shape.

Evidence that bulk fluid oscillatory movement is induced by potential gradients is provided by the fact that all organelles in the axial region of the cytoplasm move in concert and to an equal extent, not with different amplitudes as expected for an inhomogeneous population of organelles with different electrophoretic mobilities. In addition, free cytoplasmic organelles from disrupted cells fail to respond to the same electrical gradients.

It is unknown why different OHCs show different polarity of the displacement responses. Structural and metabolic differences between OHCs isolated from different turns or different rows in the cochlear spiral^{20,21} may alter the balance of forces resulting from the electro-osmotic fluid drag and the electrophoresis of soluble metabolites¹⁶ as well as of components of the membrane¹³ and of the cytoskeleton.

The requirement of video enhancement to visualize the shape changes and the intracellular organelle movements limited our

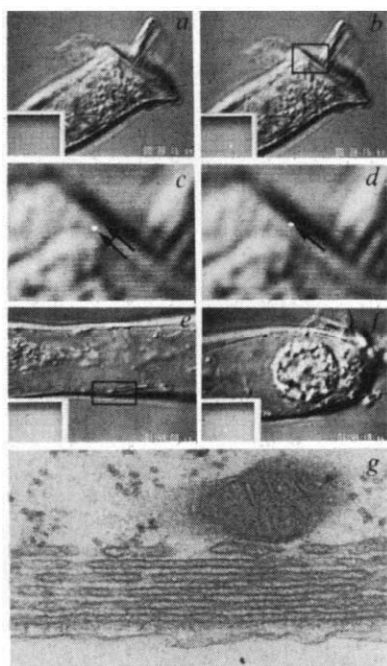


Fig. 2 Mechanical responses of OHCs to transcellular stimulation. *a*, Video-enhanced image of the stereociliar end of an OHC at the positive phase of a sinusoidal stimulus (oscilloscope screen shown at lower left); the active electrode lies outside the field of view to the right, the ground electrode is positioned near the synaptic end of the cell to the left of the field of view. The time is indicated at the lower right. *b*, Image of the same cell as in *a* during the negative phase of the stimulus. *c*, *d*, Higher magnification of the images enclosed by the rectangle at the stereociliar end of OHC shown in *a* and *b* respectively; the lower arrow in *c* indicates a reference point in the screen with fixed coordinates, the upper arrow indicates the edge of the cell image. The OHC was longer during the positive phase (*a* and *c*) of the stimulus and shorter during the negative phase (*b* and *d*). *e*, *f*, Images of the middle (*e*) and synaptic end (*f*) of another cell; cytoplasmic organelles are clearly visualized. *g*, Thin-section electron micrograph from a region of OHC equivalent to that outlined by the rectangle in *e*, showing the laminated cisternal system adjacent and parallel to the plasma membrane. This system is found between the cuticular plate and the nucleus. There are 6–8 cisternae in guinea pig OHCs; the outermost layer is parallel to and 30–35 nm from the lateral plasma membrane and each layer is parallel to its neighbours and separated from the next layer by 15–20 nm¹¹. Magnifications: *a*, *b*, *e*, *f*, $\times 1,500$; *c*, *d*, $\times 7,500$; *g*, $\times 40,000$.

ability to resolve cycle-by-cycle changes beyond 30 Hz. However, we were able to see vibration and blurring of the image of OHCs in response to electrical stimulation above 100 Hz. In addition, Ashmore and Brownell²², reproducing our experimental conditions but using a linear position detector, recorded changes in cell length in response to stimulations in excess of 8 kHz.

OHCs *in situ* have no attachments at the lateral surfaces except for synaptic contacts at the base of the cell and an apical band of tight junctions with Deiters' cells in the reticular lamina. This tight junction constrains current flow *in vivo*; its absence *in vitro* would lead to current shunting around the cell. Therefore, the actual potential gradients driving the motile process in our experiments may approach the physiological values²³ (7 V m^{-1}) measured near OHCs *in vivo*. Furthermore, smaller physiological potential gradients would produce smaller displacements which would be more compatible with observed magnitudes of vibrations of the cochlear partition—of the order of angstroms at the threshold of hearing. The free lateral surfaces of OHCs permit free expression of length changes which in turn

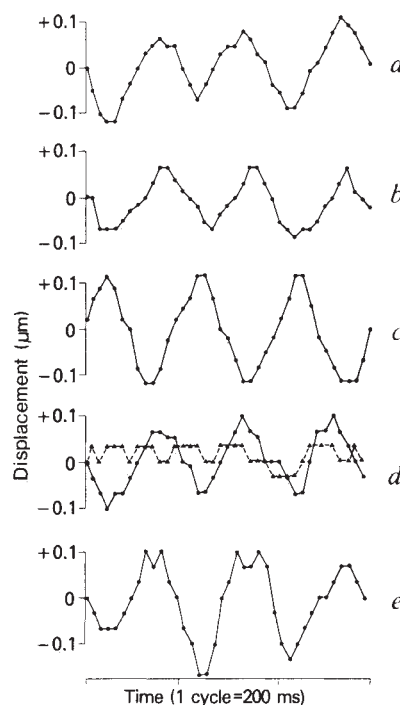


Fig. 3 Displacement evoked by 5-Hz sinusoidal transcellular stimulation. *a*–*d*, Plots of movement of the free end of OHCs; *e* is a plot of organelle movement. *a*, Apical end movement (same cell as in Fig. 2*a*–*d*). *b*, Basal end movement (same cell as in Fig. 2*f*). *c*, Example of free end movement of opposite polarity to that shown in *a* and *b*. *d*, Circles represent free end movements at reduced ionic concentration; triangles represent movements of the same cell under the same stimulus conditions, but using the original medium. We assume that a decrease in extracellular ionic strength leads to a decrease in the intracellular ionic strength; in support of this assumption are the findings of low resting potentials and input impedances of OHCs *in vitro*^{7,32}, consistent with a highly permeable plasma membrane. *e*, Organelle movement in the same cell as *d* under conditions of reduced ionic concentration; the response sinusoid is less regular because of concurrent brownian motion of the organelles.

affects the micromechanics of the cochlear partition^{7,24}. A change in the OHC length will change the separation between the reticular lamina and the basilar membrane and affect cochlear transduction^{7–10,25–27}. There is evidence that the efferent olivocochlear synapses on OHCs may modify the micromechanics of the latter^{27,28,33}. In addition, electrical potentials (cochlear microphonics)²⁹ that vary at acoustic frequencies are generated by mechano-electrical transduction associated with the OHCs³⁰, which increases the significance of an electrokinetic basis for OHC length changes by permitting shape changes in the audio-frequency ranges.

We thank Remy Pujol for the electron micrograph shown in Fig. 2*g*. This research was supported in part by research grants BNS-12174 from NSF(USA), NS-19050 from NINCDS and a research stipend from INSERM (France) to W.E.B.

Received 6 February; accepted 22 April 1986.

1. Iwasa, K., Tasaki, I. & Gibbons, R. C. *Science* **210**, 338–339 (1980).
2. Tasaki, I. & Birne, P. M. *Brain Res.* **301**, 265–272 (1984).
3. Tasaki, I. & Nakaye, T. *Science* **223**, 411–413 (1984).
4. Brown, K. T. & Murakami, M. *Nature* **201**, 626 (1964).
5. Hill, B. C., Schubert, E. D., Nokes, M. A. & Michelson, R. P. *Science* **196**, 426–428 (1977).
6. Crawford, A. C. & Fettiplace, R. *J. Physiol., Lond.* **364**, 359–379 (1985).
7. Brownell, W. E., Bader, C. R., Bertrand, D. & Ribaupierre, Y. *Science* **227**, 194–196 (1985).
8. Zenner, H. P., Zimmermann, U. & Schmitt, U. *Hearing Res.* **18**, 127–133 (1985).
9. Davis, H. *Hearing Res.* **9**, 79–90 (1983).
10. Dallos, P. in *Contemporary Sensory Neurobiology* (eds Correia, M. J. & Perachio, A. A.) 207–230 (Liss, New York, 1985).
11. Saito, K. *Cell Tissue Res.* **229**, 467–481 (1983).

12. McLaughlin, S. & Mathias, R. T. *J. gen. Physiol.* **85**, 699-728 (1985).
13. Poo, M. A. *Rev. Biophys. Bioengng* **10**, 245-276 (1981).
14. Chance, B., Williams, G. R. & Hollunger, G. *J. biol. Chem.* **278**, 439-444 (1961).
15. Epstein, M. L., Sheridan, J. D. & Johnson, R. G. *Expl Cell Res.* **104**, 25-30 (1977).
16. Horwitz, B. *Neuroscience* **12**, 887-905 (1984).
17. Balasubramanian, A. & McLaughlin, S. *Biochim. biophys. Acta* **685**, 1-5 (1982).
18. Morrison, F. A. & Osterle, J. F. *J. chem. Phys.* **43**, 2111-2115 (1965).
19. Nee, T. W. J. *Chromatography* **105**, 231-248 (1975).
20. Lim, D. J. *J. acoust. Soc. Am.* **67**, 1686-1695 (1980).
21. Bohne, B. A. & Carr, C. D. *J. acoust. Soc. Am.* **77**, 153-158 (1985).
22. Ashmore, J. & Brownell, W. E. *J. Physiol., Lond. (Abstr.)* (in the press).
23. Brownell, W. E., Manis, P. B., Zidanic, M. & Spirou, G. A. *J. Acoust. Soc. Am.* **74**, 792-800 (1983).
24. Brownell, W. E. & Kachar, B. in *Peripheral Auditory Mechanisms* (eds Allen, J. B. *et al.*) (Springer, New York, 1986).
25. Strelhoff, D., Flock, A. & Minner, K. E. *Hearing Res.* **18**, 169-175 (1985).
26. Weiss, T. F. *Hearing Res.* **7**, 353-360 (1982).
27. Brown, N. C. & Nutall, A. J. *J. Physiol., Lond.* **354**, 625-646 (1984).
28. Mountain, D. C. *Science* **210**, 71-72 (1980).
29. Wever, E. G. *Physiol. Rev.* **46**, 102-127 (1966).
30. Dallos, P. A. *Rev. Psychol.* **32**, 153-190 (1981).
31. Inoue, S. *J. Cell Biol.* **89**, 346-356 (1981).
32. Flock, A. & Strelhoff, D. *Hearing Res.* **15**, 11-18 (1984).
33. Siegel, J. H. & Kim, D. O. *Hearing Res.* **6**, 171-182 (1982).

Ionic basis of membrane potential in outer hair cells of guinea pig cochlea

J. F. Ashmore & R. W. Meech

Department of Physiology, The Medical School,
University of Bristol, University Walk, Bristol BS8 1TD, UK

Mammalian hearing involves features not found in other species, for example, the separation of sound frequencies depends on an active control of the cochlear mechanics^{1,2}. The force-generating component in the cochlea is likely to be the outer hair cell (OHC), one of the two types of sensory cell through which current is gated by mechano-electrical transducer channels sited on the apical surface³. Outer hair cells isolated *in vitro* have been shown to be motile^{4,5} and capable of generating forces at acoustic frequencies⁶. The OHC membrane is not, however, electrically tuned, as found in lower vertebrates⁷⁻⁹. Here we describe how the OHC resting potential is determined by a Ca^{2+} -activated K^+ conductance^{10,11} at the base of the cell. Two channel types with unitary sizes of 240 and 45 pS underlie this Ca^{2+} -activated K^+ conductance and we suggest that their activity is determined by a Ca^{2+} influx through the apical transducer channel, as demonstrated in other hair cells¹². This coupled system simultaneously explains the large OHC resting potentials observed *in vivo*^{13,14} and indicates how the current gated by the transducer may be maximized to generate the forces required in cochlear micromechanics.

Outer hair cells were isolated from the guinea pig cochlea and maintained *in vitro* at room temperature with no obvious change in their morphology for about 4 h (Fig. 1A; see also Fig. 1 legend). Their input conductance and resting potential, measured using whole-cell recording techniques¹⁵, were found to depend on the contents of the recording pipette. When the pipette contained 140 mM K^+ , the input conductance ranged from 25 to 40 nS (Fig. 1B, a). Typical resting potentials were -15 to -40 mV and only exceptionally were more negative. Comparable results were obtained with conventional high-resistance micropipettes (J.F.A., unpublished). Cells equilibrated with pipettes containing 140 mM Na^+ had a lower input conductance (2.9 ± 0.9 nS ($\pm$ s.d.), $n=6$; Fig. 1B, b). A low input conductance (4.8 nS) was also obtained with pipettes containing 100 mM CsCl. There was no significant difference in input conductance when the external Cl^- was replaced by gluconate. We conclude that under our experimental conditions the hair cell membrane is relatively impermeable to Cl^- and is more permeable to K^+ than to Na^+ by a factor of at least 5. Furthermore, hair cells *in vivo* have input conductances of ≥ 20 nS¹⁶; thus, the values reported here suggest that the isolation procedure introduces little change in membrane leakage.

In some experiments with 140 mM Na^+ in the recording pipette, depolarizing current pulses elicited what appeared to be a slowly developing active response (Fig. 1B, b). Such responses were not seen when the pipette contained 140 mM K^+ (Fig. 1B, a) unless it also contained the Ca^{2+} buffer BAPTA¹⁷ (Fig. 1B, c). With 10 mM BAPTA in the recording pipette, intracellular levels of calcium would be expected to fall below 10^{-8} M. Thus the observed decrease in input conductance is consistent with a membrane K^+ conductance gated by internal Ca^{2+} (ref. 18). Cells loaded with BAPTA and 140 mM K^+ from the pipette had a final mean resting potential of -60 mV ($n=7$) after equilibration (2-3 min), even though the initial resting potential was close to -10 mV. We attribute the increase in resting potential to a rise in internal K^+ and the contribution of a small (<3 nS at -60 mV) residual K^+ conductance.

When BAPTA-containing cells were studied under voltage clamp, step changes in membrane potential produced a large capacitive transient followed by a steady inward current (with hyperpolarizing pulses) or a slowly developing outward current (with depolarizing pulses; Fig. 2A). Both inward and outward currents were reduced or blocked by externally applied Cd^{2+} , which blocks Ca^{2+} currents in many different tissues¹⁹ (Fig. 2B, c).

The residual ionic currents in saline containing 1 mM Cd^{2+} appeared to be largely time-independent; they were smaller in the outward direction than in the inward direction (Fig. 2B, C) but this anomalous rectification was not characterized further. In Fig. 2D, currents recorded in Cd^{2+} have been subtracted from those recorded in normal saline to show the Cd^{2+} -sensitive fraction. The Cd^{2+} -sensitive current makes only a minor contribution to the total membrane conductance (1-4 nS at -60 mV); it appears even at membrane potentials more negative than -100 mV, that is, beyond the normal operating range for Ca^{2+} channels, and is unlikely to be the remains of the Ca^{2+} -activated K^+ conductance because it appears even in cells containing 50 mM BAPTA. We propose, instead, that it crosses the membrane via the transducer pathway which, in other vertebrate hair cells, is known to be closed by Ca^{2+} -channel blockers¹².

A prominent component of the Cd^{2+} -sensitive current is apparent during membrane depolarization as a slowly developing outward current; this appears even in cells containing 50 mM BAPTA and its tail-current, seen when the membrane is repolarized, reverses at ~ -50 mV (not shown) and so it too may be associated with the transducer channel. However, H^+ conductances appear in this voltage range in some cells and they are also blocked by Cd^{2+} (refs 20, 21). A rapidly developing Cd^{2+} -sensitive Ca^{2+} -activated K^+ current has been reported in isolated hair cells from the sacculus of the bullfrog⁷ but its time-course and voltage-dependence do not correspond to the Cd^{2+} -sensitive currents reported here.

Single-channel recordings revealed several distinct channel populations which underlie some of the conductances observed in whole cells. In the cell-attached mode²² at least two populations were evident: an uncharacterized voltage-dependent channel with a unit conductance of ~ 30 pS, and a range of larger conductance channels (40-65 pS) subsequently identified as two populations of Ca^{2+} -activated K^+ channel. Unitary K^+ currents recorded from such cell-attached membrane patches reversed direction at potentials between -20 and -30 mV, which is consistent with decreased levels of internal K^+ in isolated cells.

Current records from single Ca^{2+} -activated K^+ channels were obtained with inside-out patches of membrane²³ taken from the basolateral region of the outer hair cells. One type of channel had a unit conductance of 233 ± 33 pS ($n=10$) when bathed bilaterally in 140 mM KCl (Fig. 3A). This type of channel resembles the large-conductance Ca^{2+} -activated K^+ channel described in a variety of other systems²⁴⁻³¹ because its probability of opening increased with elevated Ca^{2+} at the internal membrane surface (Fig. 3C).

In addition to the large-conductance channel, a second class

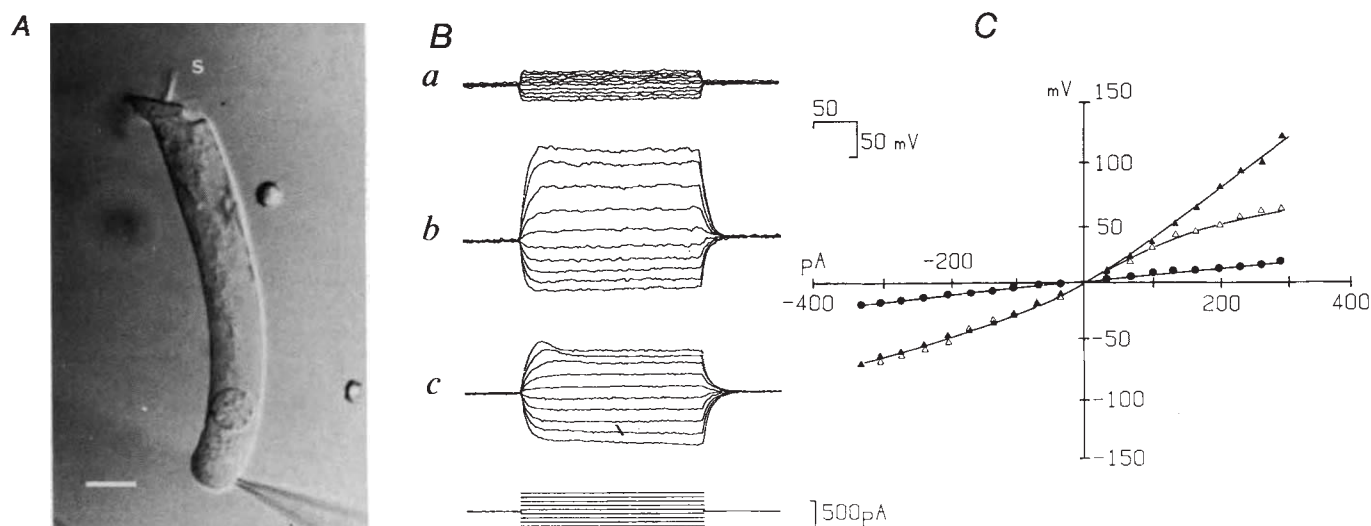
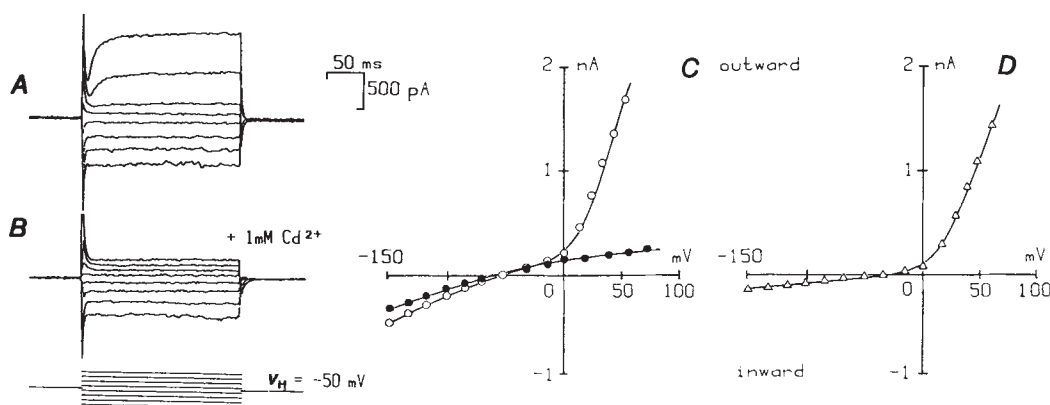


Fig. 1 Membrane permeability of isolated outer hair cells. **A**, An isolated outer hair cell of the guinea pig cochlea. The cell length indicates that the cell originates from the low-frequency end of the cochlea. The mechano-electrical transducer and stereocilia (s) on the apical surface endolymph when *in situ*. Patch recordings were made on the basal surface as shown. Scale bar, 10 μ m. **B**, Effect of different internal ions on membrane conductance of outer hair cells. The figure shows the changes in voltage observed during step changes in current, obtained in the whole-cell recording mode¹⁵ when the membrane potential had reached a steady level, that is, after 1–3 min (separate experiments with fluorescent dyes suggest that this corresponds to the time required for the cell contents to reach a steady state). The pipette contained: *a*, KCl, 140 mM, MgCl₂, 2 mM, 6.5×10^{-8} M Ca²⁺ (using 500 μ M EGTA/Ca²⁺ buffer), HEPES 5 mM; pH 7.6. Cell resting potential –29 mV. *b*, Same as in *a* except that NaCl replaced KCl. Cell resting potential +17 mV. In some cells depolarizing current pulses elicited a small, slowly developing active response on the rising phase of the trace. *c*, Same as in *a* except that 10 mM BAPTA (BDH) replaced the EGTA/Ca²⁺ buffer in the pipette. Cell resting potential –63 mV. **C**, Steady-state current/voltage curves obtained from the data in **B**. ●, Data from *a*; ▲, data from *b*; △, data from *c*. Zero on the voltage axis corresponds to the resting potential. Data were corrected for series resistance of the pipette (20–45 M Ω) in each cell.

Methods. Guinea pigs (300–500 g) were killed by rapid cervical dislocation, and their cochlear coils dissected out into L-15 medium (Gibco) using a fine needle. The apical two turns only were used; they yielded outer hair cells which were 55–80 μ m long and had an electrical capacitance in whole-cell clamp of 24–32 pF. Each coil was removed into a 140- μ l aliquot of medium containing 0.3 mg ml⁻¹ trypsin (Sigma type III S) for 30 min. The cells were mechanically dissociated and the volume increased to 300 μ l in the microscope chamber. The major ions in L-15 medium are Na⁺ (145 mM), K⁺ (5.4 mM), Mg²⁺ (1.8 mM), Ca²⁺ (1.26 mM), Cl⁻ (147 mM) and phosphate buffer. Cells were observed under 40 $\times$ WI DIC optics. Patch recordings were made from the basal part of the cell using an EPC-5 amplifier (List Electronic). Pipettes were pulled on a two-stage puller, (BB-CH, Mecanex) and used without further treatment. Seal resistances >2 G Ω were obtained before whole-cell recording. Experiments were performed at room temperature (19–21 °C).

Fig. 2 Cd²⁺-sensitive K⁺ conductance in outer hair cells recorded under voltage clamp. The pipette was filled as for Fig. 1C (10 mM BAPTA included). No cancellation of the capacitive transients was made; time constant 1.7 ms. Holding potential (V_H) = –50 mV. Step increment, 34 mV. **A**, 5 min after whole-cell recording conditions had been established. **B**, After addition of 1 mM Cd²⁺ to the bath. **C**, Current-voltage curves showing:

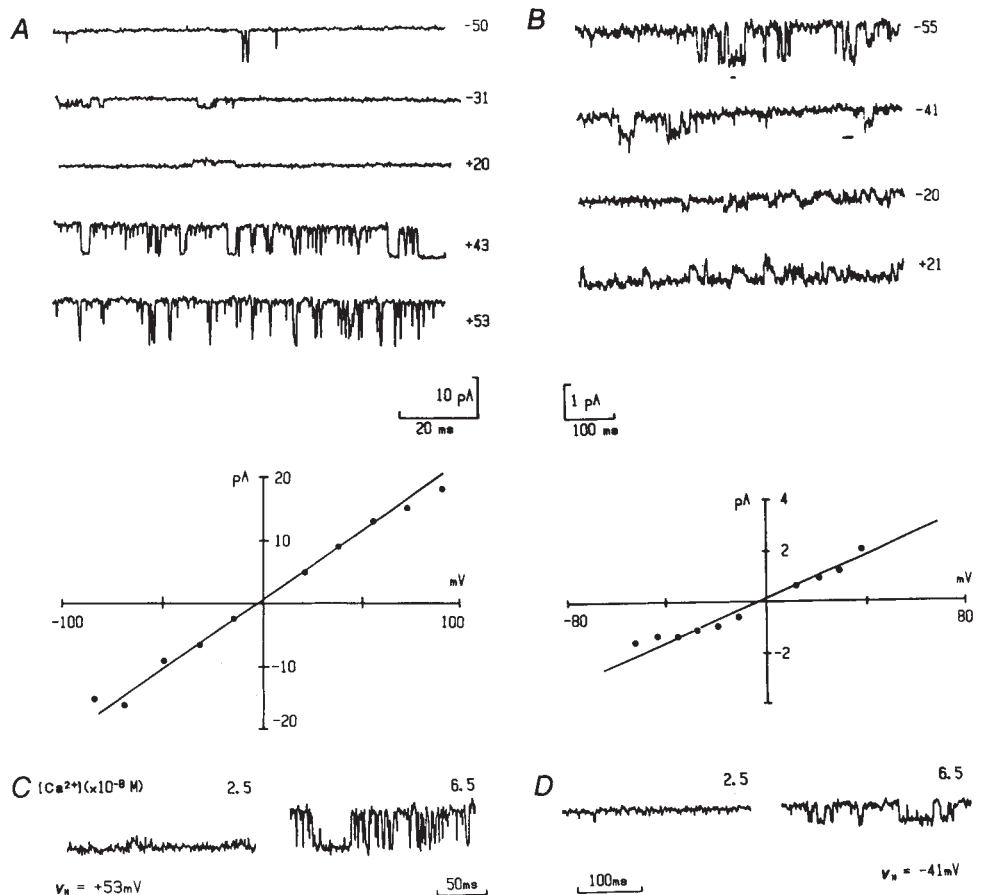


○, steady currents during depolarizing or hyperpolarizing pulses in normal saline (data from **A**); ●, steady currents in 1 mM Cd²⁺ (data from **B**). **D**, Cd²⁺-sensitive currents obtained by subtraction of Cd²⁺-resistant currents from currents observed in normal saline. The pipette series resistance (38 M Ω) measured in the bath was used to compensate **C** and **D**. Maximum slope conductance in **A** was 32 nS.

of Ca²⁺-activated K⁺ channel was found (Fig. 3B,D) which had a smaller single unit conductance, 44 ± 17 pS ($n=6$), when bathed bilaterally in 140 mM KCl. Ca²⁺-activated K⁺ channels with comparable conductances (25–92 pS in 140 mM K⁺) have been described in other vertebrate tissues^{27,32}. Like the large-conductance channel, this channel 'turned on' at Ca²⁺ levels between 2.5×10^{-8} and 6.5×10^{-8} M (with EGTA buffer, pH 7.6) but its kinetics were less complex. We have investigated the possibility that it was a fragment of some other channel. We

found that membrane patches generally had either large- or small-conductance channels and at no time did we observe large-conductance channels decomposing into smaller ones. The selectivity of the two Ca²⁺-gated channels for K⁺ over Na⁺ was about the same; the reversal potential shifted by 35–55 mV per 10-fold reduction of K⁺ at the inside surface. In these experiments we found that Na⁺ at the internal surface of the membrane produced a 'flickery' block of both types of K⁺ channel at positive membrane potentials^{31,33,34}.

Fig. 3 Ca^{2+} -activated K^+ currents in outer hair cells. Single channels recorded from inside-out patches. **A**, Large-conductance Ca^{2+} -activated K^+ channels observed at different patch potentials, as indicated; inward current downwards. The channel open probability increased with depolarization of the intracellular surface of the isolated patch (~ 30 mV for an e-fold change). The pipette contained (in mM): 140 KCl, 1.25 CaCl_2 , 2 MgCl_2 , 5 HEPES pH 7.4; the cytoplasmic face was bathed in 140 mM KCl, 6.5×10^{-8} M Ca^{2+} (500 μM EGTA/ Ca^{2+} buffer), 2 mM MgCl_2 , 5 mM HEPES pH 7.6. In bilateral 140 mM KCl, the unit slope conductance was 220 pS, measured from well-resolved openings. The K/Na permeability of this channel, determined by altering K^+ at the cytoplasmic surface, was $\sim 7:1$. Recording bandwidth (via 4-pole Bessel filter) 1.6 kHz. **B**, Small-conductance Ca^{2+} -activated K^+ channel observed under the same conditions as in **A**. Unit slope conductance 45 pS. Recording bandwidth 880 Hz. Large-conductance (**C**) and small-conductance (**D**) channel types are activated by levels of Ca^{2+} at their intracellular surface that are greater than 2.5×10^{-8} M. Values shown alongside each record show the prevailing Ca^{2+} concentration (in μM) at the cytoplasmic surface; this was changed using a continuous flow system³⁸. The holding potential V_H (+53 mV for large-conductance channel; -41 mV for small-conductance channel) was the same for both Ca^{2+} solutions. In 2.5×10^{-8} M Ca^{2+} the channels were generally inactive at voltages between -70 and +70 mV. EGTA buffers were used throughout.



In summary, we have been able to convert isolated outer hair cells with a high input conductance and low resting potential into cells with a resting potential near normal; we have done this by introducing the Ca^{2+} buffer BAPTA into the cell interior and by raising the level of internal K^+ . We have also identified two populations of channels which underlie this resting Ca^{2+} -activated K^+ conductance. *In vivo* the transducing surface of an outer hair cell faces a K^+ -rich endolymph at a potential of +80 mV; the basolateral surface faces a perilymph with the composition of normal extracellular fluid, and the cell resting potential is close to -80 mV^{13,14,35,36}. Isolated outer hair cells will tend to lose their internal K^+ because the large K^+ reservoir which drives K^+ across the apical surface *in vivo* is absent when the isolated cells are placed in normal (5.5 mM) K^+ saline. Low levels of internal K^+ and a gain of Na^+ would account for the low resting potentials of isolated cells, the reversal potential of K^+ -selective channels in the cell-attached mode and their 'flickery' block at positive membrane potentials.

Our demonstration, in isolated cells, of a Ca^{2+} -sensitive conductance at potentials more negative than -100 mV leads us to suppose that outer hair cells have a transduction channel much like that described by Ohmori for chick vestibular hair cells¹². Such channels have a high permeability for Ca^{2+} and are blocked by Ca^{2+} -channel blocking agents. Perhaps damage to the cochlea due to noise (which often involves outer hair cells) arises from a reduced Ca^{2+} influx through the disrupted transduction pathway. The reduced activity of the Ca^{2+} -activated channels that we have described would account for the previously reported decrease in K^+ permeability in noise-damaged cells³⁷.

At no time during whole-cell recording did we observe an oscillatory ringing of the membrane potential like that found in auditory hair cells of lower vertebrates⁷⁻⁹. The voltage-dependent channels we have observed were activated at voltages outside the normal range for mammalian cells^{13,14}, therefore we conclude that transduction in outer hair cells and lower vertebrate hair cells operates along different principles.

Isolated outer hair cells exhibit motile responses when current is injected into them⁴⁻⁶. This movement may be seen even in Cd^{2+} -containing solution and it appears to depend on transmembrane current rather than on the polarity of the axially flowing component. The function of the Ca^{2+} -activated K^+ conductance may be to clamp the cell close to its K^+ equilibrium potential and maximize the inward driving force for cations across the transduction channel at the apical surface. In this scheme the large outward current generated across the basolateral surface would directly couple the sensory stimulus to the cell's mechanical response.

This work was supported by the MRC, R.W.M. is a Wellcome Trust Senior Lecturer in Physiology.

Received 12 December 1985; accepted 18 April 1986.

1. Brown, M. C. & Nuttall, A. L. *J. Physiol., Lond.* **354**, 625-646 (1984).
2. Davis, H. *Hearing Res.* **9**, 79-90 (1983).
3. Pickles, J. O. *Prog. Neurobiol.* **24**, 1-42 (1985).
4. Ashmore, J. F. *J. Physiol., Lond.* **364**, 4P (1985).
5. Brownell, W. E., Bader, C. R., Bertrand, D. & de Ribaupierre, Y. *Science* **227**, 194-196 (1985).
6. Ashmore, J. F. & Brownell, W. E. *J. Physiol., Lond.* (in the press).
7. Lewis, R. S. & Hudspeth, A. J. *Nature* **304**, 538-541 (1983).
8. Ashmore, J. F. & Pitchford, S. *J. Physiol., Lond.* **364**, 39P (1985).

9. Crawford, A. C. & Fettiplace, R. *J. Physiol., Lond.* **312**, 377-412 (1981).
10. Meech, R. W. *Comp. Biochem. Physiol.* **42A**, 493-499 (1972).
11. Meech, R. W. & Standen, N. B. *J. Physiol., Lond.* **249**, 211-239 (1975).
12. Ohmori, H. *J. Physiol., Lond.* **359**, 189-218 (1985).
13. Cody, A. R. & Russell, I. *J. Nature* **315**, 662-665 (1985).
14. Dallos, P., Santos-Sacchi, J. & Flock, A. *Science* **218**, 582-584 (1982).
15. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85-100 (1981).
16. Russell, I. *J. Nobel Symp.* **63**, (in the press).
17. Tsien, R. Y. *Biochemistry* **19**, 2396-2404 (1980).
18. Meech, R. W. A. *Rev. Biophys. Bioengng* **7**, 1-18 (1978).
19. Hagiwara, S. & Byerly, L. A. *Rev. Neurosci.* **4**, 69-125 (1981).
20. Byerly, L., Meech, R. & Moody, W. Jr *J. Physiol., Lond.* **351**, 199-216 (1984).
21. Thomas, R. C. & Meech, R. W. *Nature* **299**, 826-828 (1982).
22. Sigworth, F. J. & Neher, E. *Nature* **287**, 447-449 (1980).
23. Horn, R. & Patlak, J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6930-6934 (1980).
24. Barrett, J. N., Magleby, K. L. & Pallotta, B. S. *J. Physiol., Lond.* **331**, 211-230 (1982).
25. Cook, D. L., Ikeuchi, M. & Fujimoto, W. Y. *Nature* **311**, 269-271 (1984).
26. Findlay, I. *J. Physiol., Lond.* **350**, 179-195 (1984).
27. Inoue, R., Kitamura, K. & Kuriyama, H. *Pflügers Arch. ges. Physiol.* **405**, 173-179 (1985).
28. Marty, A. *Nature* **291**, 497-500 (1981).
29. Maruyama, Y., Petersen, O. H., Flanagan, P. & Pearson, G. T. *Nature* **305**, 228-232 (1983).
30. Wong, B. S., Lecar, H. & Adler, M. *Biophys. J.* **39**, 313-317 (1982).
31. Yellen, G. *J. gen. Physiol.* **84**, 157-186 (1984).
32. Grygorczyk, R., Schwarz, W. & Passow, H. *Biophys. J.* **45**, 693-698 (1984).
33. Marty, A. *Pflügers Arch. ges. Physiol.* **396**, 179-181 (1983).
34. Singer, J. J. & Walsh, J. V. Jr *Biophys. J.* **45**, 68-70 (1984).
35. Bosher, S. K. & Warren, R. L. *Nature* **273**, 377-378 (1978).
36. Bosher, S. K. *Acta oto-lar.* **90**, 219-229 (1980).
37. Konishi, T. & Salt, A. N. *Expl Brain Res.* **40**, 457-463 (1980).
38. Yellen, G. *Nature* **296**, 357-359 (1982).

Lithium-induced respecification of pattern in *Xenopus laevis* embryos

Kenneth R. Kao, Yoshio Masui & Richard P. Elinson

Department of Zoology, University of Toronto, Toronto, Ontario, Canada M5S 1A1

Much interest in vertebrate embryology is now focused on early pattern formation in the frog, *Xenopus laevis*. In this species, the body plan is specified by a stable positional system set up by a cytoplasmic rotation in the zygote that occurs before first cleavage¹⁻⁴. Perturbation of this initial cellular event by a variety of means causes permanent distortions of the positional system⁴⁻⁷. Until now it has not been possible to alter the positional system after it has been specified. However, we report here that lithium, when applied after specification of the body plan, can respecify the positional system of the *Xenopus* embryo such that dorsal, axial structures develop from cells that otherwise contribute to ventral structures. Lithium is usually considered to have negative effects on early embryo development⁸⁻¹⁰, but our results show that lithium can act in a positive manner to produce structures which represent the uppermost values of the positional system. This discovery introduces a convenient means to study cellular and molecular mechanisms of early vertebrate pattern expression.

When dejellied 32-cell *Xenopus laevis* embryos are exposed to Li⁺ (such as 20% Steinberg's solution with 0.3 M LiCl for 6 min) they develop exaggerated head or dorso-anterior structures. They form multiple eyes or large bands of retinal pigment, and in some batches of eggs, the embryos develop a long proboscis. These embryos resemble morphologically those produced by exposure to deuterium oxide (D₂O) before axis specification, and are considered to have enhanced dorsal structures⁴. Backström⁹ reported similar results which he considered to be Li⁺-induced 'ventralization'. However, the parallels between the morphologies produced by Li⁺ and those produced by D₂O suggest that Li⁺ is causing enhancement of dorso-anterior development.

To confirm that Li⁺ causes dorso-anterior development, we applied Li⁺ to cleavage stage embryos which had been irradiated with ultraviolet (UV) light before first cleavage^{6,7}. UV-irradiated embryos lack all dorsal structures including the central nervous system, notochord and somites, and they remain as a radially symmetric ventral mass. However, dorsal structures can be

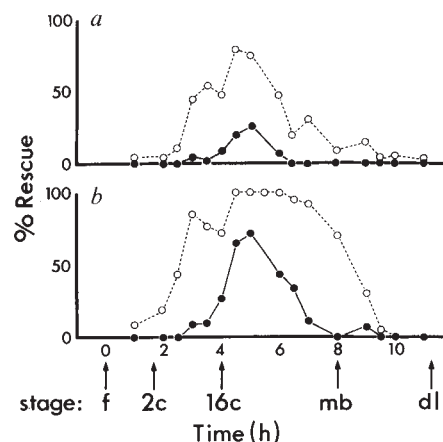


Fig. 1 Rescue of dorso-anterior development by LiCl in radially symmetric, ventral embryos produced by UV-irradiation. The horizontal axis represents the time in hours following fertilization with corresponding stages of development, f, fertilization; 2c, two cells; 16c, sixteen cells; mb, mid-blastula; dl, dorsal lip. The vertical axes represent the percentage of embryos showing axis development in response to 0.2 M Li⁺ (a) and 0.3 M Li⁺ (b) (35-87 embryos per treatment group). ○, Percentage of embryos showing dorsal development, ranging from those showing somites and tail fin (IAD grade 4) to those showing dorso-anterior radial development as in Fig. 2a and 2b. ●, Percentage of embryos showing exclusive dorso-anterior development. Note the peak in sensitivity to Li⁺ at 5 to 6 hours of development (32 to 64 cells). As judged from the area under each curve, the higher concentration of Li⁺ was able to effect more rescue than the lower concentration.

Methods. *Xenopus laevis* eggs were fertilized and dejellied as described previously¹². The vegetal poles were irradiated with UV light for 2 min at a distance of 2½ cm through a quartz slide within 30 min after fertilization. For these experiments, we used a pulse treatment of a high Li⁺ concentration for short duration in order to resolve the period of Li⁺ sensitivity. Irradiated embryos were incubated for 5 min at various times after fertilization in solutions of either 0.2 M or 0.3 M LiCl dissolved in 20% Steinberg's solution. Irradiated embryos not treated with LiCl developed into radially symmetric, ventral embryos. A few (usually <1%) develop a small tail fin but with no dorso-anterior development. After the LiCl treatment, embryos were thoroughly washed several times with 20% Steinberg's solution immediately and intermittently for several hours thereafter. All experiments and raising of embryos were done at 20 °C. Embryos were scored for development when untreated controls were at stage 40¹³.

rescued when these ventralized embryos are exposed to Li⁺ during early cleavage. Two different patterns of rescue emerge from such treatment: dorso-posterior and dorso-anterior rescue. The dorso-posterior embryo is identical to the grade-4 axis-deficient embryo classified by Scharf and Gerhart's index of axis deficiency (IAD)⁷. The grade-4 embryo has somites and a tailfin but completely lacks a head. The other type of rescued embryo develops only dorso-anterior structures. Some develop a head with a cement gland and small amount of retinal pigment (Fig. 2a), while others have radial dorso-anterior development with a wide band of retinal pigment and a cement gland encircling the head (Fig. 2b). The frequency of embryos displaying exclusive dorso-anterior rescue and those showing rescue of any dorsal structures depends on dosage of Li⁺ and the time of treatment (Fig. 1). The most sensitive period for rescuing UV-irradiated embryos with Li⁺ is between the 32- and 64-cell stages. Thus, the presence of dorsal structures in embryos otherwise lacking them suggests that Li⁺ is activating, some time during early cleavage, those components that are responsible for dorsal differentiation.

The radial arrangement of dorso-anterior structures produced by Li⁺ treatment suggests that Li⁺ is acting evenly around the embryo. We microinjected Li⁺ into specific cells of UV-irradi-

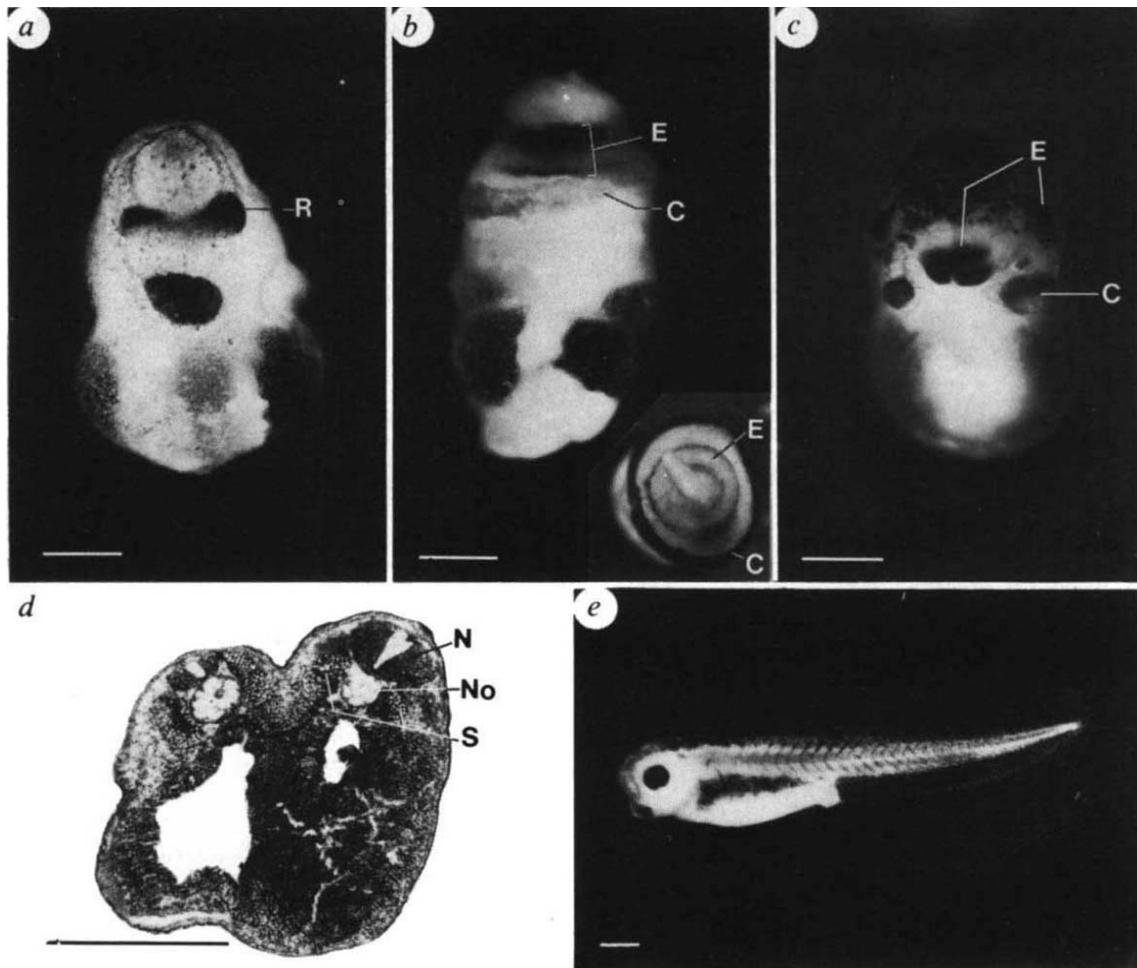


Fig. 2 *a*, *X. laevis* embryo exposed to UV light prior to first cleavage and then treated with 0.3 M LiCl for 5 min at the 32-cell stage. Note the presence of retinal pigment (R) which would not develop without LiCl exposure. *b*, An embryo with dorso-anterior radial symmetry. This embryo, like that in *a*, was UV-irradiated. Exposure to 0.3 M LiCl for 5 min 4.5 h following fertilization (16–32 cells) restored dorso-anterior development, causing rescue of eye (E) and cement gland (C) in a radial manner (inset: top view of radial, anterior embryo). *c*, A 'Janus twin' embryo exhibiting duplication of dorso-anterior structures. This embryo was injected with 4 nl of 0.3 M Li⁺ into a ventral, vegetal cell at the 32-cell stage. The eyes (E) and cement gland (C) of the head on the right side are labelled. *d*, A section just posterior to the head of an embryo displaying duplication of the head with one posterior axis. This embryo was microinjected similarly to the embryo in *c*. Embryos were fixed, sectioned in paraplast and stained as described previously¹². The duplicate axis is unlabelled. No, notochord; N, neural tube; S, somites. See Fig. 4 for experimental details. *e*, A control, untreated embryo at stage 40. Scale bars, 0.5 mm.

ated embryos to try to restore dorso-anterior development to one side of the embryo and thereby rescue normal development. Embryos at the 32-cell stage were used because they could be easily injected without significant leakage, and were most sensitive to rescue by Li⁺. Microinjection of Li⁺ into a cell in the vegetal-most tier caused significant rescue of normal development in a dose-dependent manner (Fig. 3). We used the IAD to quantify axis development, because the embryos developed morphologies identical to those defined by the IAD scale. For example, in Fig. 3, vegetal cell injections of 0.3 M Li⁺ reduced the average IAD to 2 in embryos which would otherwise develop into grade 4.9 or 5 embryos. Out of 81 embryos, 30% were scored as grade 0 (normal), 13% as grade 1 (slightly microcephalic), 16% as grade 2 (cyclopic), 17% as grade 3 (microcephalic), 16% as grade 4 (acephalic), and 8% as grade 5 (radial ventral). We could not assign an IAD score to embryos injected with solutions of Li⁺ greater than 0.3 M. These concentrations approach lethal levels and in the surviving embryos, there is an over-enhancement of dorso-anterior structures so that the embryos tend towards radial dorso-anteriorization. Injection of Li⁺ into an animal cell gave a lower frequency of rescue than with vegetal cell injections. It remains to be determined whether

the injected vegetal cell is altered by Li⁺ or whether it acts as a local source of Li⁺.

The restoration of bilateral symmetry in UV-irradiated embryos by Li⁺ microinjection prompted us to determine whether dorsal structures are duplicated in normal embryos when they are microinjected with Li⁺. We microinjected into vegetal lower-tier cells as this injection gave the best rescue of UV-irradiated embryos. From 112 embryos microinjected with Li⁺ into a ventral, vegetal cell, we obtained 97 embryos that had duplication of dorso-anterior structures (Fig. 4). Most of these embryos developed into 'Janus' twins (Fig. 2c) that lack posterior development but consist of two heads which have normally formed eyes and cement glands. A small number (~5%) developed into twins which have a single posterior axis but with twinned heads, central nervous systems, notochords and somites (Fig. 2d). Injection into a dorsal cell failed to duplicate dorsal structures. Out of 97 embryos injected into a dorsal, vegetal cell, 89 survived to form a single body axis. Most (~80%) developed normally, but the remainder developed enhanced dorso-anterior structures, similar in external appearance to the 'imbalanced' embryos produced by Cooke⁵.

To see whether dorsal rescue is specific for Li⁺, we microinjec-

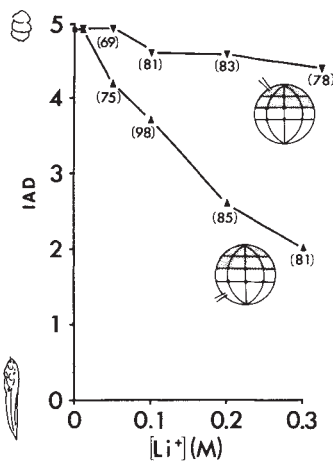


Fig. 3 Rescue of normal development by Li^+ microinjection into radially symmetric ventral, embryos. Embryos were fertilized and irradiated with UV light as described in Fig. 1, and then microinjected with 3–5 nl of LiCl dissolved in distilled H_2O or 200% Steinberg's solution at varying concentrations (horizontal axis) into either a top most animal tier cell ($\blacktriangledown$) or lowest vegetal tier cell ($\blacktriangle$). The embryos for each experimental group were scored for axis development (vertical axis) using the index of axis deficiency (IAD)⁷. An IAD score of 0 indicates normal development whereas a score of 5 indicates complete lack of dorsal structures with only radial ventral development. The average IAD is plotted here for each group of injected embryos. Note that the average IAD is reduced in both animal and vegetal cell injections, but more significantly in the latter. Microinjection of $>0.3 \text{ M Li}^+$ increased the frequency of dead embryos and the formation of exclusively dorso-anterior embryos that are not scorable with the IAD. The numbers in brackets refer to numbers of embryos injected with LiCl at the indicated concentration. (For 0.01 M LiCl , 78 embryos were microinjected in the animal cell and 86 embryos were injected into the vegetal cell. At 0 M , 269 control embryos were scored.)

ted 4 nl of 0.4 M solutions of NaCl , KCl , CsCl , RbCl or NH_4Cl , 50 mM solutions of CaCl_2 or MgCl_2 , or a 10 mM solution of ZnCl_2 (higher concentrations of the latter three salts usually killed the embryos). All solutions were made up in 200% Steinberg's solution which, when injected alone into UV-irradiated embryos, did not cause rescue. Among the cations tested, only Li^+ was able to cause significant rescue of dorsal structures.

Based on cell transplant experiments Gimlich and Gerhart¹¹ showed that a vegetal, dorsal cell at the 32-cell stage carried sufficient information to promote complete dorsal development. This information is translated by inductive cell interactions during cleavage to form dorsal elements. Our experiments show that Li^+ is able to cause expression of dorsalizing information in cells that otherwise lack or do not express this information. Thus, the potential for cells to undergo dorsal development exists radially around the embryo even after axis specification, and such development will occur when stimulated by Li^+ at the appropriate time.

The ability to produce radial dorsal embryos with Li^+ should facilitate investigations on the molecular and cellular differences between dorsal and ventral pattern formation in *X. laevis*.

We thank JoAnn Render for initial discussions. This work was supported by a grant to R.P.E. from NSERC, Canada.

Received 12 March; accepted 25 April 1986.

1. Elinson, R. P. *Symp. Soc. dev. Biol.* **38**, 217–234 (1980).
2. Gerhart, J. C., Ubbels, G., Black, S., Hara, K. & Kirschner, M. *Nature* **292**, 511–516 (1981).
3. Vincent, J.-P., Oster, G. F. & Gerhart, J. C. *Dev. Biol.* **113**, 484–500 (1986).
4. Scharf, S. R., Vincent, J.-P. & Gerhart, J. C. *UCLA Symp. molec. cell. Biol.* **19**, 51–73 (1984).
5. Cooke, J. *Nature* **319**, 60–63 (1986).
6. Grant, P. & Wacaster, J. F. *Dev. Biol.* **28**, 454–471 (1972).
7. Scharf, S. R. & Gerhart, J. C. *Dev. Biol.* **99**, 75–87 (1983).
8. Lehmann, F. *Einführung in die physiologische Embryologie* (Birkhäuser, Basel, 1945).
9. Backström, S. *Archiv. Zool.* **6**, 527–536 (1954).
10. Masui, Y. *Experientia* **17**, 1–6 (1961).
11. Gimlich, R. & Gerhart, J. C. *Dev. Biol.* **104**, 117–130 (1984).
12. Kao, K. R. & Elinson, R. P. *Dev. Biol.* **107**, 239–251 (1985).
13. Nieuwkoop, P. D. & Faber, J. *Normal Table of Xenopus laevis (Daudin)* (Elsevier, Amsterdam, 1967).
14. Kirschner, M. W. & Hara, K. *Mikroskopie* **36**, 12–15 (1980).

Oral vaccination of the fox against rabies using a live recombinant vaccinia virus

J. Blancou*, M. P. Kieny†, R. Lathe‡, J. P. Lecocq†, P. P. Pastoret‡, J. P. Soulebot§ & P. Desmettre§

* Ministère de l'Agriculture, Centre National d'Etudes sur la Rage et la Pathologie des Animaux Sauvages, BP 9, 54220 Malzeville, France

† Transgène SA, 11 rue de Molsheim, 67000 Strasbourg, France

‡ Faculté de Médecine Vétérinaire, Université de Liège,

45 rue des Vétérinaires, 1070 Bruxelles, Belgique

§ Rhône-Mérieux, Laboratoire IFFA, 254 rue Marcel Mérieux, 69342 Lyon, France

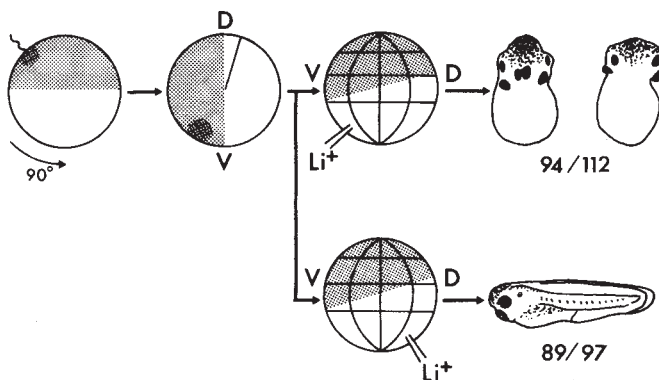


Fig. 4 Summary of twinning experiments. Fertilized eggs were placed in a 5% solution of Ficoll (Pharmacia) in 20% Steinberg's solution and rotated within 30 min after fertilization with the sperm entry point (SEP) towards gravity. The rotation ensures the position of the dorsal side of the embryo (D) as 180° opposite the SEP (ventral side, V) in the meridian passing through the SEP and the animal-vegetal axis. Eggs were kept in this orientation for 20 to 30 min and the dorsal (upper) side was marked with Nile blue sulphate¹⁴. After the rotation period, eggs were allowed to orient normally in 20% Steinberg's solution with the vegetal pole downwards. At the 32-cell stage, embryos were microinjected either in a ventral, vegetal-most cell or a dorsal, vegetal-most cell with 3–5 nl of 0.3 M LiCl dissolved in 200% Steinberg's solution. At stage 40, the embryos were scored for development either as having a single body axis or with duplicated dorso-anterior development. The twins resulting from ventral cell injection are usually Janus twins with two diametrically opposed anterior ends, although some show closely approximated heads as in Fig. 2c.

Rabies, a viral disease affecting all warm-blooded animals, is prevalent in most parts of the world¹, where it propagates amongst wild animals, particularly the fox and dog. The public health and economic consequences of infection in man and livestock are well known. Attempts to control the disease by vaccinating wild carnivores with inactivated or attenuated rabies virus remain controversial, and we have instead evaluated here the potential of a recombinant vaccinia virus to protect foxes against the disease. We have found that the administration of vaccinia virus (VV) or a recombinant harbouring the rabies surface antigen gene (VVTGgRAB) is innocuous to foxes. The recombinant virus can elicit the production of titres of rabies-neutralizing antibodies equal or superior to those obtained with conventional vaccine, and 10^8 plaque-forming units (PFU) of VVTGgRAB administered subcutaneously, intradermally or orally confers complete protection to severe challenge infection with street rabies virus.

Table 1 Rabies-neutralizing antibodies and resistance to challenge

Vaccine and route	Dose (PFU)	Animals	Rabies neutralizing antibody titre*	Mean titre	Resistance to challenge†	Fraction surviving
No vaccine		#	0	NA	—	0/6
Conventional vaccine subcutaneous‡		440	1.21	1.49	+	2/2
		441	1.77		+	
VVTGgRAB, intradermal§	10 ⁸	446	3.03	2.82	+	2/2
		449	2.61		+	
VVTGgRAB, subcutaneous§	10 ⁸	439	3.03	NA	+	2/2
		442	0		+	
VVTGgRAB, oral scarified	10 ⁸	437	1.91	2.4	+	4/4
		438	2.33		+	
		447	2.61		+	
		452	2.75		+	
VVTGgRAB, oral	10 ⁴	408	1.35	NA	+	1/4
		427	0		—	
		428	0		—	
		429	0		—	
	10 ⁶	416	0.8	0.4	+	2/4
		425	0.8		+	
		426	0		—	
		430	0		—	
	10 ⁸	414	2.33	2.57	+	4/4
		431	2.61		+	
		433	2.61		+	
		451	2.75		+	
VVTGgRAB in bait	10 ⁸	411	1.07	1.8	—¶	4/5
		412	1.63		+	
		420	2.61		+	
		423	2.19		+	
		424	1.49		—¶	

Rabies neutralizing antibody titres were determined in accordance with recommendations laid down by the World Health Organisation²⁷. Titres are expressed as the log₁₀ of the final neutralizing dilution (FND). For conversion to international units (IU) IU = 59/[antilog (3.5 - log FND)]. Foxes were monitored for 50 days following challenge and the presence of rabies virus in the brain of the animals succumbing to challenge was verified by immunofluorescence (not shown). Challenge virus GS6 (street virus comprising salivary glands from rabid foxes dispersed in phosphate-buffered saline (PBS) (20% w/v) as described previously²⁸) was injected in a volume of 1 ml containing 5,700 mouse LD₅₀ units (intracerebral), ~17,000 fox LD₅₀ units (intramuscular). Vaccinia Copenhagen strain and VVgRAB were grown on tissue culture cells as described elsewhere²⁹. Supernatants were recovered, the cell pellets homogenized, re-centrifuged and pooled with the supernatant. Virus was purified by sedimentation through a sucrose cushion (36% w/v) for 2 h, 14,000 r.p.m., Beckman SW28 rotor. Pellets were suspended into PBS, sonicated briefly and banded by centrifugation on a sucrose gradient (20–40% w/v; 12,000 r.p.m., 45 min, SW28) before diluting (PBS) to the required concentration. Virus was titred on BHK21 cells. NA, not applicable.

* Only the 28-day titre is given.

† +, Resisted challenge, —, succumbed to rabies. All animals not resisting died between 15 and 25 days after challenge.

‡ 'Rabisin ND' vaccine, lot 5532; neutralizing antibody titres obtained with live attenuated virus are similar (not shown).

§ Virus was inoculated intradermally as described (see text); subcutaneous injections were performed in a volume of 1 ml.

|| Virus was administered by direct application into the mouth by syringe (vol. 1 ml).

¶ Two animals were observed to have ingested only a part of the vaccine.

Four animals were vaccinated with vaccinia virus Copenhagen strain, two animals received no treatment.

Transmission of rabies occurs predominantly through biting. Large quantities of virus are shed in the saliva and the aggressive behaviour of the rabid animal facilitates transmission. Infection by this route is invariably fatal and an immune population is therefore never established. Prophylactic measures have aimed at eliminating or reducing the population of the principal reservoir through poisoning or gassing², though a decrease in the carrier population density rarely reduces the rate of advance of the disease front³. In Europe, culling of foxes has reduced the number of rabies outbreaks but has not contained the disease. Vaccination, perhaps supported by culling in low-density fox populations, seems more likely to be effective than mere destruction of foxes⁴. An immune population is better able to limit the spread of the disease than one which is sparse but susceptible^{4,5}.

Oral administration could facilitate the vaccination of large numbers of wild foxes and this was first attempted in North America and Europe^{6,7}. Live attenuated rabies virus introduced into chicken heads, sausages or dog biscuits and distributed in the wild has been successfully used to vaccinate wild foxes⁸, and field trials are in progress in Switzerland, West Germany⁹ and Canada (C. D. MacInnes, personal communication).

However, the virus is often unstable and attenuated viruses retain pathogenicity for rodents and can revert to virulence¹⁰. Furthermore, inactivated rabies virus is ineffective when administered orally¹¹. A novel vaccination strategy, in which a recombinant vaccinia virus bearing a foreign antigen coding sequence is used as the immunizing agent^{12–14}, seemed to hold some promise for the vaccination of foxes against rabies.

The causative agent of rabies is a rhabdovirus, and the glycoprotein (G) which traverses the envelope surrounding the virus, is the sole viral protein capable of inducing and reacting with virus-neutralizing antibodies or of conferring protection against rabies^{15,16}. The relative innocuity of vaccinia virus, which has been used extensively to control and eradicate smallpox in man¹⁷, has stimulated its development as a cloning and expression vector, and derivatives expressing surface antigens from influenza, hepatitis B and herpes simplex have been used to confer protection against these diseases¹⁴. We recently developed a recombinant VV (VVTGgRAB) expressing the rabies G coding sequence¹⁸ and found that mice scarified with the live recombinant virus VVTGgRAB resisted severe challenge with street rabies virus^{18–20}. VV itself is an enveloped virus and

the recombinant VVTGgRAB presenting rabies G protein elicited protection against rabies even after chemical inactivation¹⁹. We therefore extended our investigations to wild foxes.

Vaccinia virus (Copenhagen strain) was first tested for innocuity to foxes. European foxes (*Vulpes vulpes*) of both sexes captured in the wild and raised in captivity as described previously²¹ were inoculated with live vaccinia virus. Two animals (nos 445, 448) received VV by injection (vol. 0.1 ml) into the depilated skin of the back and two further animals (446, 449) received the recombinant VVTGgRAB by the same route. 10^2 , 10^4 , 10^6 or 10^8 PFU of virus were injected at triplicate sites and cutaneous reactions monitored for 15 days.

Mild localized inflammation was observed at the sites of injection with 10^6 or 10^8 PFU of virus, and in one animal (445) at 10^2 and 10^4 PFU. In all cases inflammation regressed spontaneously within 8 days. Cutaneous reaction was significantly greater with the wild-type strain of VV than with the recombinant VVTGgRAB strain (not shown). No lesions appeared elsewhere than at the site of injection and there was no evidence here of contact transmission of vaccinia virus between animals. In test animals receiving live recombinant vaccinia virus by direct application into the mouth (Table 1) no impairment of digestive or alimentary function was observed.

We subsequently examined the ability of the recombinant virus to elicit the production of antibodies directed against rabies virus. Animals inoculated intradermally, subcutaneously or orally with VVTGgRAB were bled on days 0, 8, 14 and 28; serum was separated and titred for the presence of rabies-neutralizing antibodies. High titres of neutralizing antibodies were present in sera from all but one animal treated with 10^8 PFU of VVTGgRAB (Table 1). Further, scarification of oral mucous membranes before oral administration (to facilitate penetration of the virus) did not yield improved titres of neutralizing antibodies. All animals presented only low levels of antibody capable of neutralizing VV (not shown), in agreement with other reports²².

Direct protection testing was then performed. Twenty-eight days after vaccination, animals were challenged by injection of 5,700 mouse LD₅₀ (50% lethal dose) units of rabies virus. All 12 animals receiving 10^8 PFU of VVTGgRAB either orally or parenterally resisted challenge (Table 1). One animal (442) exhibiting undetectable levels of rabies-neutralizing antibodies also resisted challenge, attesting to the relevance of cell-mediated immunity in defence against rabies^{16,23}. Control animals receiving no vaccine succumbed to the disease after 15–25 days. Two animals injected subcutaneously with a commercial inactivated (adjuvanted) vaccine similarly resisted challenge, although virus-neutralizing antibodies were present at a reduced level (Table 1). Oral administration of conventional (inactivated) vaccine has previously been shown to be ineffective¹¹.

Animals receiving less than 10^8 PFU of VVTGgRAB showed a clear dose-dependent response, with one out of four and two out of four animals surviving challenge after oral administration of 10^4 and 10^6 PFU of VVTGgRAB, respectively (Table 1).

Oral administration is the only route appropriate to the vaccination of wild animals. Accordingly, the vaccine must be presented in a form suitable for ingestion. We thus prepared 'Plas-tipak' capsules (a gift from Dr Wandeler) containing 10^8 PFU of VVTGgRAB, inserted them into chicken heads (one capsule per head into the beak) and distributed them to test animals (one per fox). These animals similarly produced high titres of rabies-neutralizing antibodies and resisted severe challenge with rabies virus (Table 1).

Transmissibility is a major factor determining the impact of a live vaccine on a wild population. We therefore sought to establish whether animals vaccinated with VVTGgRAB might transmit the virus by contact with untreated control animals. Four test animals (two male, two female) were acclimatized (4 days) to sharing a pen (2 m²) with an animal of the opposite sex. One animal from each pair received 10^8 PFU of live

VVTGgRAB by direct application into the mouth (vol. 1 ml). Serum samples were collected from both vaccinated and control animals at 14 and 28 days, and intramuscular challenge with live rabies virus (see Table 1 legend) was performed at 28 days.

All vaccinated animals presented high titres of rabies-neutralizing antibodies (mean 2.72 at 28 days, units as in Table 1) and resisted challenge. In three out of four control animals no rabies-neutralizing antibodies were detectable and these animals succumbed to challenge (not shown). Surprisingly, the fourth control animal (523, female) presented significant levels of rabies-neutralizing antibodies (1.35 and 0.97 at 14 and 28 days, respectively) and resisted severe challenge infection.

Subsequent investigations into possible mechanisms of transmission revealed that both the relevant vaccinated male (517) and control female 523 consistently displayed aggressive behaviour, and reciprocal biting was observed within a few minutes of oral vaccination. Contamination of bite wounds with VVTGgRAB may, in this instance, have been sufficient to effect immunization. Such a mode of transmission requires a combination of rather exceptional circumstances, and we surmise that contact transmission of VVTGgRAB in the field is likely to be rare.

We have shown that vaccinia virus and its recombinant VVTGgRAB bearing the rabies surface antigen are innocuous to healthy foxes; indeed, the thymidine kinase-negative recombinant may be more innocuous than wild-type strains of vaccinia virus²⁴. Our recombinant virus can confer complete protection to severe challenge infection with rabies virus. Similar experiments are in progress with other animal vectors of rabies, notably the skunk and racoon²⁵. Importantly, presentation to foxes of the live recombinant, encapsulated and introduced into chicken heads, also yields animals resisting severe challenge infection. Procedures for the production, stabilization and distribution of vaccinia virus are fully established²⁶. Due to its efficacy, innocuity and stability, the recombinant virus may be a candidate for the large-scale vaccination of wild foxes and, possibly, other feral or domestic animals.

We thank H. Koprowski and T. Wiktor for helpful discussion and for communicating unpublished data, R. Thomas for his practical suggestions concerning this study, D. Schmitt, K. Dott, J. Bailly, M. Slevé and E. Cain for technical assistance and N. Monfrini and L. Schneider for help in preparing the manuscript. We dedicate this manuscript to the late Dr T. Wiktor.

Received 10 March; accepted 18 April 1986.

- Steck, F. *Symp. zool. Soc. Lond.* **50**, 57–75 (1982).
- Debbie, J. G. in *Report on Rabies*, 22–27 (Veterinary Learning Systems, Princeton, 1983).
- Toma, B. & Andral, L. *Adv. Virus Res.* **21**, 1–36 (1977).
- Anderson, R. M., Jackson, H. C., May, R. M. & Smith, A. D. M. *Nature* **289**, 765–771 (1981).
- Andral, L. & Blancou, J. *Rev. Sci. Tech. Off. int. Epiz.* **1**, 927–959 (1982).
- Baer, G. M., Ableseth, M. K. & Debbie, J. G. *Am. J. Epidemiol.* **93**, 487–490 (1971).
- Mayr, A., Kraft, H., Haeger, O. & Haake, H. *Zentbl. VetMed.* **19**, 615–625 (1972).
- Blancou, J. *Rec. Méd. Vét.* **155**, 733–741 (1979).
- Schneider, L. G., Cox, J. H. & Muller, W. W. *Rev. Ecol. (Terre Vie)* **40**, 265–266 (1985).
- Pépin, M. *et al. Annu. Inst. Pasteur Virol.* **136E**, 65–73 (1985).
- Blancou, J., Andral, L., Aubert, M. F. A. & Artois, M. *Bull. Acad. vét. Fr.* **55**, 351–359 (1982).
- Panicali, D., Davis, S. W., Weinberg, R. L. & Paoletti, E. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5364–5368 (1983).
- Smith, G. L., Mackett, M. & Moss, B. *Nature* **302**, 490–495 (1985).
- Smith, G. L., Mackett, M. & Moss, B. *Biotechnol. genet. Engng Rev.* **2**, 383–407 (1984).
- Wunner, W. H., Dietzschold, B., Curtis, P. & Wiktor, T. J. *J. gen. Virol.* **64**, 1649–1656 (1983).
- Kieny, M. P., Desmetre, P., Soulebot, J. P. & Lathé, R. *Prog. vet. Microbiol. Immun.* **3** (in the press).
- Behbehani, A. M. *Microbiol. Rev.* **47**, 455–509 (1983).
- Kieny, M. P. *et al. Nature* **312**, 163–166 (1984).
- Wiktor, T. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 7194–7198 (1984).
- Lathé, R. *et al. in Vaccines 85* (eds Lerner, R., Chanock, R. & Brown, F.) 157–162 (Cold Spring Harbor Laboratory, New York, 1985).
- Dubreuil, M., Andral, L., Aubert, M. F. A. & Blancou, J. *Ann. Rech. Vét.* **10**, 9–21 (1979).
- Appleyard, G. & Andrews, C. J. *gen. Virol.* **23**, 197–200 (1974).
- Wiktor, T. J. *Dev. Biol. Standard.* **40**, 225–264 (1978).
- Buller, R. M., Smith, G. L., Moss, B., Cremer, K. & Notkins, A. L. in *Vaccines 85* (eds Lerner, R., Chanock, R. & Brown, F.) 163–167 (Cold Spring Harbor Laboratory, New York, 1985).
- Wiktor, T. J. *Inst. Pasteur Symp. Vaccines and Vaccination* (in the press).
- World Health Organisation *Tech. Rep. Ser.* **323** (WHO, Geneva, 1966).
- Kaplan, M. M. & Koprowski, H. *La Rage: Techniques de Laboratoire* 3rd edn (World Health Organisation, Geneva, 1974).
- Blancou, J., Aubert, M. F. A., Andral, L. & Artois, M. *Rev. Méd. vét.* **130**, 1001–1015 (1979).
- Drillien, R., Tripiet, F., Koehren, F. & Kim, A. *Virology* **79**, 369–380 (1977).

Preferential expression of a defined T-cell receptor β -chain gene in hapten-specific cytotoxic T-cell clones

Ute Hochgeschwender*†, Hans Ulrich Weltzien†, Klaus Eichmann†, R. Bruce Wallace* & Jörg T. Epplen*

* Junior Research Unit and † Department of Cellular Immunology, Max-Planck-Institut für Immunbiologie, D-7800 Freiburg, FRG

A multitude of different antigens can be recognized by T cells through specific receptors. Both the α - and β -chains of the T-cell receptor contribute to the antigen recognition portion¹. The repertoire of β -chain variable region (V_β) gene segments is limited to some 20 elements²⁻⁴ which seem to be used randomly in different T cells^{2,5,6}. Diversity at the β -chain level can be created in several ways^{2,3,7-12}: a multiplicity of germline gene segments; combinatorial diversity by rearranging different V, diversity (D), joining (J) and constant (C) region elements; junctional diversity by joining gene segments at different sites; N-region diversity, that is, insertion of random nucleotides at junctional sites; and somatic mutation¹³. However, the major sources and the extent of diversity of the T-cell receptor are unclear. To address this issue, 42 H-2K^b-restricted, 2,4,6-trinitrophenyl (TNP)-specific cytotoxic T-cell (T_c) clones from C57BL/6 mice were characterized with respect to expression of different β -chain gene segments in messenger RNA using specific oligonucleotide probes. We report here that nearly half of the T_c clones use identical elements for productive β -chain gene rearrangement. Thus, there is a restriction in the use of β -chain gene segments in this panel of T_c clones which favours a particular V_β - D_β - J_β - C_β combination with a defined D_β element.

A collection of 42 class I (H-2K^b) restricted cytotoxic T-cell clones possessing specificity for the hapten TNP has been established¹⁴. In order to isolate and characterize the expressed T-cell receptor β -chain gene of a randomly chosen T-cell clone, a complementary DNA library was constructed in the bacteriophage vector λ gt11 from total cellular RNA of T_c clone 112-2. The nucleotide sequence of a β -chain cDNA clone (pBJ2.1) was determined as $V_\beta 3$ - $D_\beta 112$ -2- $J_\beta 2$ -6- $C_\beta 2$ (nomenclature according to ref. 2; $V_\beta 3$ is termed $V_\beta 2B4$ in ref. 7 and $V_\beta 13$ in ref. 3), whereby $D_\beta 112$ -2 is a segment of seven nucleotides partly different from the known germline sequences of $D_\beta 1$ -1 or $D_\beta 2$ -1 (Fig. 1a).

Synthetic oligonucleotides were prepared which were specific for the V, D and J segments of T_c clone 112-2 (Fig. 1a) in order to determine the frequencies with which the $V_\beta 3$, $D_\beta 112$ -2 and $J_\beta 2$ -6 gene segments are used in the panel of T_c clones. Furthermore, oligonucleotides specific for the $C_\beta 1$, $C_\beta 2$ and C_α gene elements were used for sequential Northern blot or direct gel hybridizations. An example is shown in Fig. 1b in which hybridizations of an RNA gel with the set of oligonucleotide probes reveal the individual elements expressed in β -chain mRNAs of the different T_c clones. We found that 19 out of 42 (45%) T_c clones use the same $V_\beta 3$ gene segment as T_c clone 112-2 (see Fig. 2). Moreover, whenever the $V_\beta 3$ gene segment is expressed, it is always rearranged to $D_\beta 112$ -2, $J_\beta 2$ -6 and $C_\beta 2$ as in T_c clone 112-2 (see Figs 1b, 2). Therefore, 45% of the H-2K^b/TNP-specific T_c clones must use identically composed β -chain genes for the T-cell receptors. In half of the remaining T_c clones, different V_β and D_β elements are joined to the $J_\beta 2$ -6- $C_\beta 2$ gene segments, while 30% of all T_c clones use additional V_β , D_β and J_β elements in connection with either $C_\beta 1$ or $C_\beta 2$.

To obtain additional evidence that the D elements used in individual T_c clones are identical—in particular with regard to the V-D and D-J junctions—probes were designed with one

or two base mismatches at the borders of the $D_\beta 112$ -2 element (see Fig. 3). EcoRI-digested DNA of cDNA clone pBJ2.1 was probed with the perfectly matched $D_\beta 112$ -2, the D_I (one mismatch at the D-J junction), the D_{II} (two mismatches at the D-J junction) and the D_{III} (one mismatch at the V-D junction) probes under various conditions (Fig. 3). Only the perfectly matched $D_\beta 112$ -2 probe was able to form stable hybrids under the standard conditions employed. Corresponding results were observed in RNA hybridization experiments (data not shown). Therefore, the finding that 45% of T_c clones give positive signals with the $D_\beta 112$ -2 probe indicates that very similar or identical D elements are being expressed.

The $V_\beta 3$, $D_\beta 112$ -2 and $J_\beta 2$ -6 oligonucleotide probes were also employed in the DNA hybridization analysis of the panel of H-2K^b/TNP-specific T_c clones using different restriction enzymes. As an example, Fig. 4 shows the gel hybridization pattern obtained with EcoRI-digested DNA from three T_c clones and B6 liver (Fig. 4). Using the $V_\beta 3$ probe, two germline gene segments were detected in B6 liver DNA, while the band of high relative molecular mass (M_r) corresponds to the $V_\beta 3$ band of the B10.A mouse (refs 2, 12 and our unpublished results). Sequence analysis of the coding region of the B6 germline $V_\beta 3$ element revealed identity with the known segment from B10.A mice. Southern blot and DNA gel hybridizations of all $V_\beta 3$ -positive T_c clones revealed identical productive $V_\beta 3$ rearrangements (data not shown), suggesting that only the element of the high- M_r band is used. With the $V_\beta 3$ and $J_\beta 2$ -6 probes, in addition to germline elements, rearranged gene segments were detected in T_c clones. Hybridization of the $D_\beta 112$ -2 oligonucleotide to the DNAs produced a signal exclusively in a T_c clone in the same position as the productively rearranged $V_\beta 3$ and $J_\beta 2$ -6 elements (Fig. 4). The hybridization specificity and design of the $D_\beta 112$ -2 probe (Fig. 1a) ensures that a positive signal can be obtained only when $D_\beta 112$ -2 is rearranged to the $V_\beta 3$ and $J_\beta 2$ -6 gene segments. All the T_c clones that scored positive for the complete preferentially expressed V_β gene showed the single $D_\beta 112$ -2 band of Fig. 4 (data not shown). In additional Southern blot hybridization, the rearrangements of the $C_\beta 1$ and $C_\beta 2$ loci were investigated in T_c clones which did not hybridize with $V_\beta 3$, $D_\beta 112$ -2 or $J_\beta 2$ -6 probes. Nick-translated and oligonucleotide probes specific for the $J_\beta 1$ and $J_\beta 2$ clusters revealed heterogeneous rearrangements in $J_\beta 1$ for the $C_\beta 1$ -positive T_c clones and in $J_\beta 2$ for the $C_\beta 2$ -positive T_c clones (data not shown). With regard to the productive rearrangements, this heterogeneity stands in contrast to the uniformity observed in the group of T_c clones that express the defined V_β gene.

Although there have been no similar studies of T-cell receptor β -chain diversity in an antigen-specific restricted immune response, data are available on Southern hybridization experiments of cytochrome c-specific T-helper (T_h) hybridomas¹⁵ and bulk T-cell lines specific for hen egg white lysozyme⁵. Both reports present evidence that similar or identical β -chain gene rearrangements have occurred in the DNA of their isolated T_h cells from the same immune response. It is assumed, therefore, that these T_h cells express identical V_β genes. Our investigation shows clearly that in a collection of T_c clones specific for the same antigen (H-2K^b/TNP), a defined complete T-cell receptor β -chain gene is preferentially expressed. The reasons for this are unknown, but the situation is reminiscent of observations of humoral immune responses; in the mouse the antibody response to certain model antigens¹⁶⁻¹⁹ is characterized by the dominant expression of one or a few related immunoglobulin heavy-chain V-region genes by antibody-secreting lymphocytes. In contrast to the similarities to V-gene expression in B- and T-cell receptors in defined immune responses, the choice of the D segment appears to be restricted only in the β -chain of T_c clones. D-region identity is not found in specific humoral immune response. In immunoglobulins, the V regions of heavy chains are often almost identical, whereas the D segments are extremely variable¹⁶⁻¹⁸, although antigen recognition is

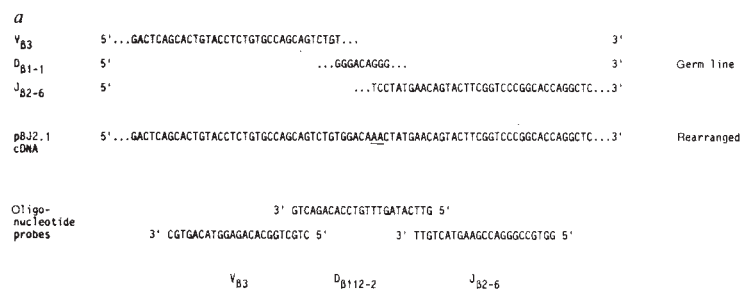
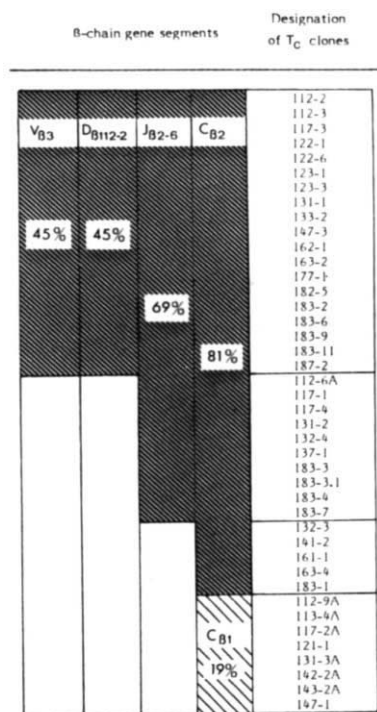
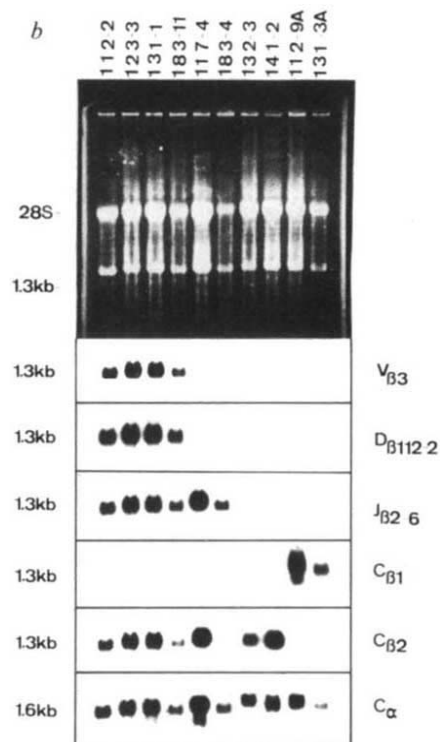


Fig. 1 a, Sequences of the V_{β} , D_{β} and J_{β} gene segments in germline configuration and after rearrangement in the TNP-specific T_c clone 112-2, illustrating junctional and N -region diversity. cDNA clone pBJ2.1 is compared with published germline gene segments $V_{\beta}3$ (ref. 7), $D_{\beta}1-1$ (ref. 11) and $J_{\beta}2-6$ (ref. 7). At least two nucleotides (underlined) at the D - J joining of pBJ2.1 are encoded by neither the germline D_{β} nor the J_{β} segment. The oligonucleotide probes $V_{\beta}3$, $D_{\beta}112-2$ and $J_{\beta}2-6$ are complementary to the mRNA and depicted below the cDNA sequence. **b**, Gel hybridization analysis of RNA from 10 different T_c clones specific for H-2K^b/TNP using six different ³²P-labelled oligonucleotide probes specific for different T-cell receptor gene segments. kb, kilobases.

Methods. a, A cDNA library was constructed from total cellular RNA of T_c clone 112-2 in the bacteriophage vector λ gt11. Screening was carried out according to ref. 22, using a nick-translated probe (AK1)²³ which hybridizes with the constant region of the T-cell receptor β -chain gene. Positive clones were isolated and plaque purified. A hybridizing cDNA clone from the 112-2 library was subcloned into pUC12 (pBJ2.1) and into M13mp9. Sequence analysis of pBJ2.1 was carried out according to Sanger *et al.*²⁴. The sequence of the 740-base-pair (bp) insert of pBJ2.1 consists of the 3' end of a V_{β} (~100 bp), a D_{β} and a J_{β} element as well as a complete C_{β} (including some 50 bp of 3' untranslated sequence). **b**, RNA from cultured cells was extracted in the presence of guanidinium thiocyanate²⁵ and pelleted by centrifugation through a 5.7 M caesium chloride cushion. RNA (10 μ g) was denatured in the presence of formaldehyde and agarose gel electrophoresis was performed²⁶. The ethidium bromide-stained gel was photographed under ultraviolet light (top) and dried²⁷ or blotted onto GeneScreen (NEN). Oligonucleotides were synthesized on an automated DNA synthesizer and radiolabelled using [γ -³²P]ATP and T4 polynucleotide kinase²⁸. All hybridizations and washing steps were carried out according to ref. 18. **c**, The same gel has been hybridized consecutively with various probes (in the order given) at 60 °C and was washed for 1 min at the indicated temperatures: The washed gel was exposed to Kodak XAR-5 films using Quanta III (Dupont-Cronex) intensifying screens for approximately 40 h. After each exposure, probes were removed by washing the gel in 5 mM EDTA at 60 °C. Washing efficiency was monitored by control exposures.



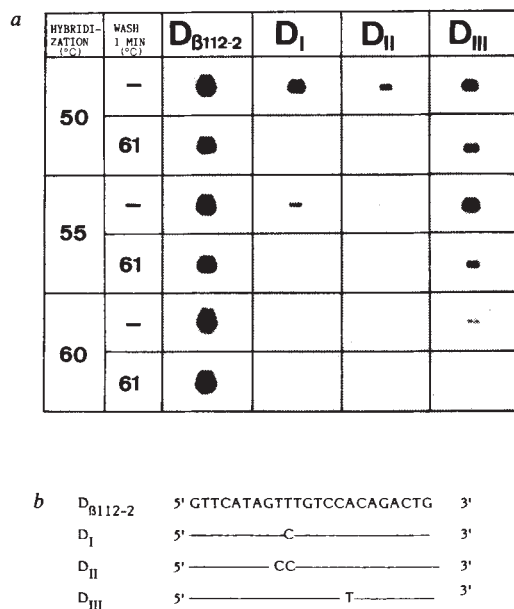


Fig. 3 *a*, Hybridization specificity of the oligonucleotide probe D_{B112-2} and three derivatives with base mismatches at the V-D and D-J junctions. pBJ2.1 (3 µg) was digested with *Eco*RI, divided into six identical aliquots and subsequently subjected to electrophoresis in a 0.75% agarose gel³¹. After denaturation, neutralization and drying of the gel²⁷, the six individual lanes were cut. *b*, Gels strips were consecutively hybridized with the D_{B112-2} probe and its derivatives. After hybridization at the indicated temperatures, each gel strip was washed in 6×SSC at room temperature. Half of the gel strips were then washed for 1 min at 61 °C as indicated. Steps were exposed for 15 h without intensifying screen. After each hybridization the probe was removed and gel strips re-exposed to exclude any remaining signal.

maintained. This indicates that the D regions of heavy chains are involved in idiotyp^{20,21}. There is no obviously analogous role for the D region in the T-cell receptor β-chain.

The finding that 45% of the H-2K^b/TNP-specific T_c clones express very similar or identical D_β elements is unexpected in view of the fact that the D_{B112-2} element is not identical to known germline D_β elements. If the germline D_{B1-1} element is assumed to be used, at least two bases (AA) at the borders of the D_β-J_β junction are not germline encoded (see Fig. 1*a*) and consequently are generated by a random process². If one assumes that the β-chain D region is important for antigen

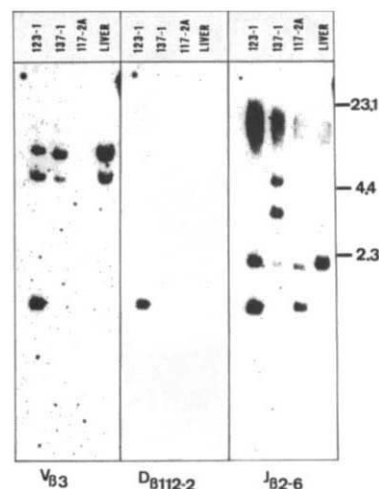


Fig. 4 Analysis of DNA from T_c clones specific for H-2K^b/TNP and from liver of the B6 mouse.

Methods. Cellular DNA was extracted modified according to ref. 31. DNA (10 µg) prepared from T_c clones was digested to completion with the restriction enzyme *Eco*RI. DNAs were electrophoresed in 0.75% agarose gels³¹. After denaturation and neutralization the gel was dried²⁷ and hybridized consecutively with the oligonucleotide probes J_β2-6, D_{B112-2} and V_β3. (See Fig. 1*b* legend for a description of the probes.) The gel was exposed for 4 days with intensifying screens. M_r markers are indicated in kilobases.

binding and that the dinucleotide AA of the AAC codon for asparagine is critical for TNP recognition, then its presence in nearly half of the T_c clones would be due to selection by the antigen. On the other hand, the D_{B112-2} element might be encoded by a hitherto undescribed D_β segment in which the preferential combination of this D element with V_β3 and J_β2-6 could also be explained by selective forces. We have found evidence (from additional oligonucleotides covering the V_β3-D_{B112-2} junction) for independence of the V_β3 and D_{B112-2} elements; it remains to be seen how the D_{B112-2} element is in fact generated.

We thank Hannelore Thoma and Ulrike Pflugfelder for tissue culture work, Gary McMaster for λgt11 DNA, Lore Lay for art work, and Rosemary Schneider for secretarial assistance. R.B.W. was recipient of a research award from the Alexander von Humboldt-Stiftung. This work was partially supported by the Deutsche Forschungsgemeinschaft. R.B.W. was on sabbatical leave from Beckman Research Institute of the City of Hope, Duarte, California 91010, USA.

Received 16 January; accepted 24 April 1986.

- Yagüe, J. *et al. Cell* **42**, 81-87 (1985).
- Barth, R. K. *et al. Nature* **316**, 517-523 (1985).
- Behlke, M. A. *et al. Science* **229**, 566-570 (1985).
- Davis, M. M. *A. Rev. Immun.* **3**, 537-560 (1985).
- Goverman, J. *et al. Cell* **40**, 859-867 (1985).
- Rupp, F., Acha-Orbea, H., Hengartner, H., Zinkernagel, R. & Joho, R. *Nature* **315**, 425-427 (1985).
- Chien, Y., Gascoigne, N. R. J., Lee, N., Kavalier, J. & Davis, M. M. *Nature* **309**, 322-327 (1984).
- Malissen, M. *et al. Cell* **37**, 1101-1110 (1984).
- Gascoigne, N. R. J., Chien, Y. H., Becker, D. M., Kavalier, J. & Davis, M. M. *Nature* **310**, 387-391 (1984).
- Kavalier, J., Davis, M. M. & Chien, Y.-H. *Nature* **310**, 421-423 (1984).
- Siu, G. *et al. Nature* **311**, 344-350 (1984).
- Patten, P. *et al. Nature* **312**, 40-46 (1984).
- Augustin, A. A. & Sim, G. K. *Cell* **39**, 5-12 (1984).
- Weltzien, H. U., Kempkes, B., Jankovic, D. L. & Eichmann, K. *Eur. J. Immun.* **16**, 631-639 (1986).
- Hedrick, S. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 531-535 (1985).

- Schilling, J., Clevinger, B., Davie, J. W. & Hood, L. *Nature* **183**, 35-40 (1980).
- Gearhart, P. J., Johnson, N. D., Douglas, R. & Hood, L. *Nature* **291**, 29-34 (1981).
- Kaartinen, M., Griffiths, G. M., Markham, A. F. & Milstein, C. *Nature* **304**, 320-324 (1983).
- Manser, T., Huang, S.-Y. & Gefter, M. L. *Science* **226**, 1283-1288 (1984).
- Gridley, T., Margolis, M. N. & Gefter, M. J. *Immun.* **134**, 1236-1244 (1985).
- Radbruch, A. *et al. Nature* **315**, 506-508 (1985).
- Benton, W. D. & Davis, R. W. *Science* **196**, 180-182 (1977).
- Rinaldi, A., Wallace, R. B., Simon, M. M., Becker, A. & Epplen, J. T. *Immunogenetics* **21**, 403-406 (1985).
- Sanger, F., Coulson, A. R., Barell, B. G., Smith, A. J. H. & Roe, B. J. *molec. Biol.* **143**, 161-178 (1980).
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294-5299 (1979).
- Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201-5205 (1980).
- Tsao, S. G. S., Brunk, C. F. & Pearlman, R. E. *Analyt. Biochem.* **131**, 365-372 (1983).
- Miyada, C. G., Reyes, A. A., Studencki, A. B. & Wallace, R. B. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2890-2894 (1985).
- Shearer, G. M. *Eur. J. Immun.* **4**, 527-533 (1974).
- Fujiwara, H. & Shearer, G. A. *J. Immun.* **126**, 1047-1051 (1981).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).

Alternative splicing of murine T-cell receptor β -chain transcripts

Mark A. Behlke & Dennis Y. Loh

Howard Hughes Medical Institute, Departments of Medicine, Microbiology, and Immunology, Washington University School of Medicine, St Louis, Missouri, 63110, USA

Variable processing of heteronuclear RNA into multiple, defined species of messenger RNA is a well-established phenomenon in a number of systems, including myelin basic protein^{1,2}, calcitonin³, troponin T⁴ and immunoglobulins⁵⁻⁷. Within a single B cell, immunoglobulin heavy-chain peptides often exist as two related but distinct species, a membrane-bound form and a secreted form, both of which originate from the same germline constant(C)-region gene via alternative RNA processing pathways⁵. Furthermore, at certain stages of B-cell development, a single variable(V)-region element can be concurrently expressed in immunoglobulins of both the IgM and IgD isotypes, presumably resulting from the alternative splicing of one variable-region exon to different constant-region genes^{6,7}. We report here the existence of an alternative RNA splicing pathway available to murine T-cell receptor (TCR) β -chain transcripts. Sequence analysis of β -chain complementary DNA clones reveals a $C_{\beta}1$ species containing a 72-base pair (bp) insertion between the joining (J_{β}) and C_{β} elements. This sequence is inserted via an alternative splicing pathway available to $C_{\beta}1$ transcripts. The optional exon is located between the $J_{\beta}1$ cluster and the first exon of $C_{\beta}1$. Interestingly, this element can be spliced to $C_{\beta}2$ in the New Zealand White mouse, in which the $C_{\beta}1$ gene is deleted^{8,9}. Use of the alternative splicing pathway varies between 1% and 18% of total C_{β} clones, depending on the source of isolation.

We previously reported the nucleotide sequences of a series of murine TCR β -chain variable-region genes identified from 15 V_{β} - C_{β} -containing cDNA clones¹⁰. All 15 of these clones contain a somatically rearranged V_{β} - D_{β} - J_{β} element and a C_{β} element, of which 13 clones have the V_{β} - D_{β} - J_{β} element spliced in the expected manner to the C_{β} element. The remaining two clones were both found to contain an identical 72-bp insertion between their J_{β} and C_{β} elements. The insertion of this 72-bp sequence maintains the open translational reading frame seen in conventionally spliced J_{β} - C_{β} -containing cDNAs and results in the addition of 24 extra codons precisely at the J_{β} - C_{β} splice junction^{11,12}. As such, the 72-bp insertion appears to define the existence of a new exon which is available for alternative splicing of β -chain transcripts. We call this new exon $C_{\beta}0$.

Of the two clones initially found to include this novel sequence, clone B/C-61 contains $V_{\beta}10$ - $D_{\beta}1.1$ - $J_{\beta}1.6$ - $C_{\beta}0$ - $C_{\beta}1$ and clone B/C-77 contains $V_{\beta}8.2$ - $D_{\beta}1.1$ - $J_{\beta}1.6$ - $C_{\beta}0$ - $C_{\beta}1$. (V_{β} nomenclature follows that of Barth *et al.*¹³, as described in Behlke *et al.*¹⁴). Subsequently, a third cDNA clone (B/C-04) which also contained the $C_{\beta}0$ exon was obtained. This clone contains $J_{\beta}1.3$ - $C_{\beta}0$ - $C_{\beta}1$ and is truncated within the J_{β} element. The nucleotide sequences of these three cDNA clones are presented in Fig. 1a. These clones all originated from a BALB/c thymus cDNA library; two out of four V_{β} -containing clones examined from this library¹⁰ contained the alternatively spliced $C_{\beta}0$ exon. To determine whether alternatively spliced $C_{\beta}0$ - C_{β} -containing clones could be detected in lymphoid tissue other than thymus, the two oligonucleotide probes described in Fig. 1c were used to screen an unamplified B6 spleen cDNA library. Four $C_{\beta}0$ -containing clones were identified from a total of 301 that were C_{β} positive. Three of these four clones contain V_{β} elements; their DNA sequences are presented in Fig. 1b. Clone B6-26s contains $V_{\beta}12$ - $D_{\beta}1.1$ - $J_{\beta}1.1$ - $C_{\beta}1$, clone B6-34 contains $V_{\beta}6$ - $D_{\beta}1.1$ - $J_{\beta}1.1$ - $C_{\beta}1$, and clone B6-73 contains $V_{\beta}1$ - $D_{\beta}1.1$ -

Table 1 Relative expression of the $C_{\beta}0$ exon in cDNA libraries

cDNA library	$C_{\beta}0$	C_{β}	Frequency of $C_{\beta}0$	Frequency of $C_{\beta}1$	Total no. of clones screened
C57BL/6 spleen	4	301	1.3%	27%	350,000
C57BL/6 thymus	14	267	5%	32%	60,000
NZW thymus	26	146	18%	*	40,000

cDNA libraries were constructed and screened as described in the legends to Figs 1 and 3. Libraries were also screened with nick-translated probes specific for $C_{\beta}1$ and $C_{\beta}2$; these probes contain only 3'-untranslated (UT) sequence and were subcloned from $C_{\beta}1$ - and $C_{\beta}2$ -containing cDNA clones. It is not clear whether the frequency of $C_{\beta}0$ -containing clones in unamplified cDNA libraries reflects the actual frequency of these species *in vivo*. While we have no way of directly addressing this question for $C_{\beta}0$ -containing transcripts, we can make such a comparison for a different subset of β -chain transcripts. The C57BL/6 spleen cDNA library has been screened for V_{β} usage (complete data to be presented elsewhere); out of 154 C_{β} -positive clones examined, 102 hybridize to probes specific for the 16 known murine V_{β} subfamilies¹⁴, including 19 clones which hybridize to $V_{\beta}8.2$. Thus, 19% (19/102) of known V_{β} -containing clones and 12% (19/154) of total C_{β} -containing clones contain $V_{\beta}8$ -subfamily genes in our B6 spleen cDNA library. The anti-TCR monoclonal antibodies F23.1 and KJ16-133 recognize 19% and 13%, respectively, of peripheral T cells in B6 mice^{30,31}. These reagents are believed to identify a subset of β -chain peptides expressing elements of the three-member $V_{\beta}8$ subfamily^{14,32}. Thus, expression of at least one subset of β -chain clones in our cDNA library approximates the known level of expression of this same subset *in vivo*. We presume the same will be true for $C_{\beta}0$ -containing clones.

* Does not apply.

$J_{\beta}1.4$ - $C_{\beta}1$. Clone B6-27 (not shown) contains $C_{\beta}0$ spliced as expected to $C_{\beta}1$, but does not contain an identifiable J_{β} or V_{β} element 5' to $C_{\beta}0$. The seven $C_{\beta}0$ -containing clones described above use five different V_{β} elements and four different J_{β} elements, thus the alternative splicing pathway does not prefer any particular J_{β} or V_{β} sequence.

All seven $C_{\beta}0$ -containing clones obtained thus far also contain $C_{\beta}1$, implying that the $C_{\beta}0$ exon should be located between the $J_{\beta}1$ cluster and $C_{\beta}1$ in the genome. A BALB/c genomic library was screened with a C_{β} probe, and a phage was isolated which contained both the $J_{\beta}1$ cluster and $C_{\beta}1$. As expected, $C_{\beta}0$ -specific probes hybridize to this genomic clone. A restriction map of the $J_{\beta}1$ - $C_{\beta}1$ region of this phage is presented in Fig. 2a, and the nucleotide sequence of the $C_{\beta}0$ -hybridizing region is presented in Fig. 2b. The nucleotide sequence of the $C_{\beta}0$ exon and its flanking introns closely match the consensus donor and acceptor splice sites established by Breathnach and Chambon¹⁵ (Fig. 2c), and the splice sites predicted by such an analysis are in complete agreement with the actual nucleotide sequences of the spliced $C_{\beta}0$ -containing cDNAs in Fig. 1. We conclude that the inclusion of the $C_{\beta}0$ exon in transcripts represents a naturally occurring alternative splicing pathway and is not an aberrant event or cloning artefact.

Thus far, $C_{\beta}0$ -containing clones have only been identified in cDNA libraries constructed from heterogeneous populations of T cells, such as from thymus or spleen, and not in libraries constructed from cloned T cells. From such data we cannot determine whether the 1.3% incidence of $C_{\beta}0$ expression seen in C57BL/6 spleen reflects a low level of alternatively spliced message in all $C_{\beta}1$ -expressing cells or if a particular subset of $C_{\beta}1$ -expressing cells selectively express $C_{\beta}0$ at higher levels. Two cDNA libraries we examined from cloned T cells both proved to have their V_{β} elements productively rearranged to $J_{\beta}2$ - $C_{\beta}2$ (ref. 10), and no $C_{\beta}0$ - or $C_{\beta}1$ -containing clones were obtained. Molecular analysis of several $C_{\beta}1$ -expressing T-cell clones have been reported^{13,16-18}, but no $C_{\beta}0$ -containing cDNAs were described in these reports. A thorough examination of

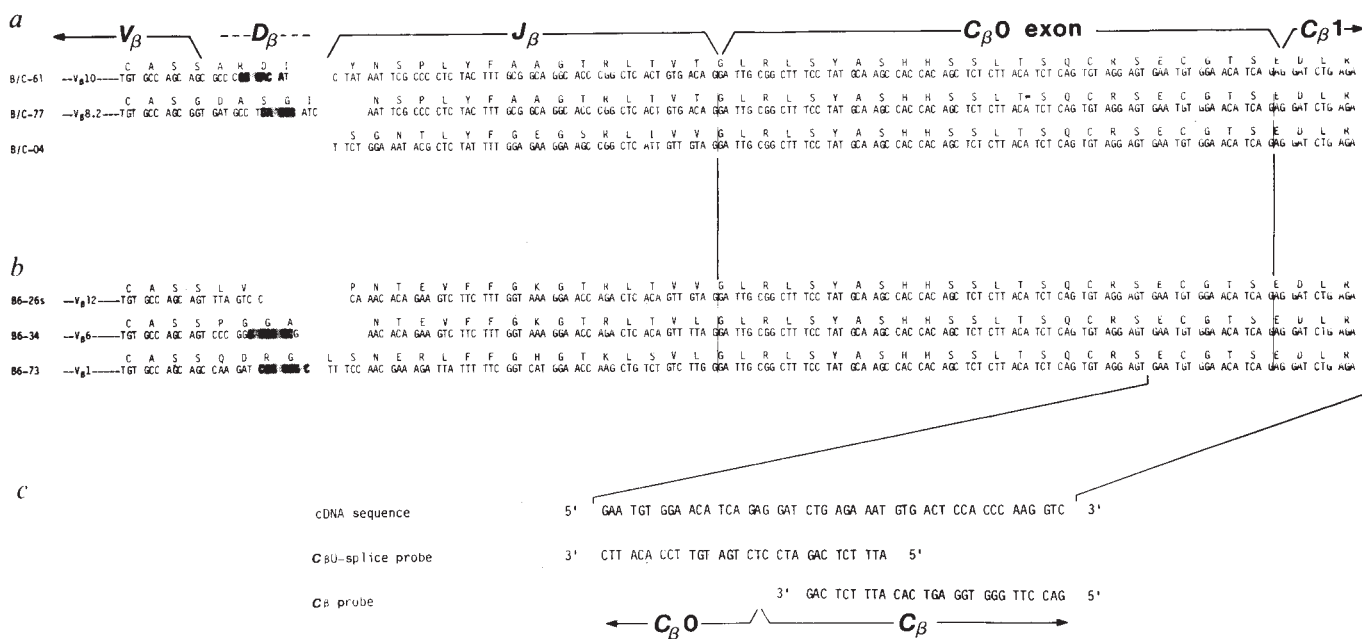


Fig. 1 Nucleotide and amino-acid (one-letter code) sequence of alternatively spliced TCR β -chain cDNAs. **a**, Isolates from BALB/c thymus. Clones containing a 72-bp insertion between the J_β and C_β elements are presented. Nucleotide and predicted amino-acid sequences begin at the 3' conserved cysteine codon of the V_β element and extend into the 5' end of C_β . Additional V_β or C_β sequence is not included as these sequences are well established^{10-14,19}; V_β gene segments are identified according to the nomenclature of Barth *et al.*¹³, as described in Behlke *et al.*¹⁴. Sequences contributed from the germline $D_\beta 1.1$ element are shaded while J_β gene segments are aligned separately. The J_β and C_β splice junctions are demarcated, and the 72-bp insertion has been labelled $C_\beta 0$. **b**, Isolates of three V_β -containing alternatively spliced cDNAs from C57BL/6 spleen. **c**, Design of C_β and $C_\beta 0$ -specific synthetic oligonucleotide probes. The nucleotide sequences of alternatively spliced cDNA clones at the $C_\beta 0$ - C_β splice junction are given, under which are aligned the sequences of two synthetic oligonucleotides which are complementary to the original cDNA sequence. The $C_\beta 0$ -splice probe spans the $C_\beta 0$ - C_β splice junction and is specific for clones in which the $C_\beta 0$ exon has been spliced to C_β . The C_β probe is complementary only to actual C_β coding sequence and hybridizes to all C_β -containing clones which extend sufficiently 5' to contain this sequence.

Methods. **a**, A BALB/c thymus cDNA library was screened with a TCR C_β -specific probe¹⁰. Random clones were subcloned into pUC12 and their nucleotide sequences were determined by the method of Maxam and Gilbert²⁵. **b**, A C57BL/6 spleen concanavalin A blast cDNA library (previously described¹⁰) was screened with a $C_\beta 0$ -specific oligonucleotide probe, as described below. $C_\beta 0$ -containing cDNAs were subcloned into pUC12 and their nucleotide sequences were determined by the method of Maxam and Gilbert²⁵. ³²P-labelled oligonucleotide probes were hybridized at 68 °C in 0.9 M NaCl, 0.18 M Tris pH 7.4, 0.012 M EDTA, 2× Denhardt's solution, and 20 $\mu\text{g ml}^{-1}$ salmon sperm DNA; washing was performed at 55 °C in 0.75 M NaCl, 0.075 M sodium citrate (5× SSC)²⁶.

several $C_\beta 1$ -expressing T-cell clones for $C_\beta 0$ expression will be required to resolve this point.

The absence of $C_\beta 0$ -like alternative splicing products in $C_\beta 2$ -containing transcripts can be explained in several ways. First, it may be that a $C_\beta 0$ analogue ($C_\beta 0'$) exists between the $J_\beta 2$ cluster and $C_\beta 2$ but that this element has diverged far enough from the $C_\beta 0$ sequence described here so that our synthetic oligonucleotide probes do not cross-hybridize with it. Our first three $C_\beta 0$ - $C_\beta 1$ -containing clones were identified from an analysis of randomly isolated C_β -positive thymus cDNA clones (Fig. 1a). All subsequent $C_\beta 0$ -containing clones were identified by hybridization to an oligonucleotide probe, which would select against the isolation of $C_\beta 0'$ - $C_\beta 2$ -containing clones if this probe did not cross-hybridize to the putative $C_\beta 0'$ sequence. A small number of $C_\beta 2$ -containing clones have been examined for this possibility, and no $C_\beta 0'$ - $C_\beta 2$ clones were found. While this may suggest that variable RNA splicing occurs exclusively in $C_\beta 1$ transcripts, it is also possible that an insufficient number of $C_\beta 2$ clones has been examined. However, we note that both nick-translated and oligonucleotide probes specific for $C_\beta 0$ do not hybridize to murine $C_\beta 2$ -containing genomic clones (data not shown), and that no sequence with an open reading frame and >50% homology to $C_\beta 0$ can be found within 900 bp 5' of the first exon of $C_\beta 2$ (ref. 19).

It has been reported that the New Zealand White (NZW) mouse contains a germline deletion which includes both $C_\beta 1$ and the $J_\beta 2$ cluster⁸, suggesting that a $J_\beta 1$ - $C_\beta 2$ -containing species should account for all NZW β -chain transcripts. The

deletion of $C_\beta 1$ in the NZW mouse is thought to have resulted from a homologous recombination event between the first exon of $C_\beta 1$ and the first exon of $C_\beta 2$ (ref. 9). The sequence 5' to the single C_β coding region should therefore be of $C_\beta 1$ origin, implying that the $C_\beta 0$ exon will be present in NZW. To look for the existence of a $C_\beta 0$ - $C_\beta 2$ -containing species and examine $C_\beta 0$ expression in the altered chromosomal context of the NZW mouse, an unamplified NZW thymus cDNA library was screened with the C_β and $C_\beta 0$ -splice oligonucleotide probes. Of 146 C_β -positive clones identified, 26 hybridized to the $C_\beta 0$ -splice probe. Thus, 18% (26/146) of the C_β -containing clones in the NZW thymus library use the $C_\beta 0$ alternative splicing pathway. This frequency is significantly higher than the 1.3% level seen earlier in the C57BL/6 spleen cDNA library.

Two of these 26 $C_\beta 0$ -containing clones were randomly chosen for further analysis. The DNA sequences of these two clones were determined and are presented in Fig. 3. Clone NZW 8 is truncated within V_β coding sequence but was found to contain $V_\beta 12$ - $D_\beta 1.1$ - $J_\beta 1.1$ - $C_\beta 0$ - $C_\beta 2$; clone NZW 22s is full-length and contains $V_\beta 5.2$ - $D_\beta 1.1$ - $J_\beta 1.4$ - $C_\beta 0$ - $C_\beta 2$. While the $V_\beta 12$ sequence seen in clone NZW 8 is identical to a V_β sequence reported previously¹⁰, the V_β sequence in clone NZW 22s has not been reported; we note, however, that this sequence is 88% homologous at the nucleotide level to the published $V_\beta 5.1$ sequence¹³ and is therefore presumed to represent the second member of a three-member V_β gene subfamily, whose existence was inferred from Southern blot analysis using the $V_\beta 5.1$ sequence as a probe^{13,14}.

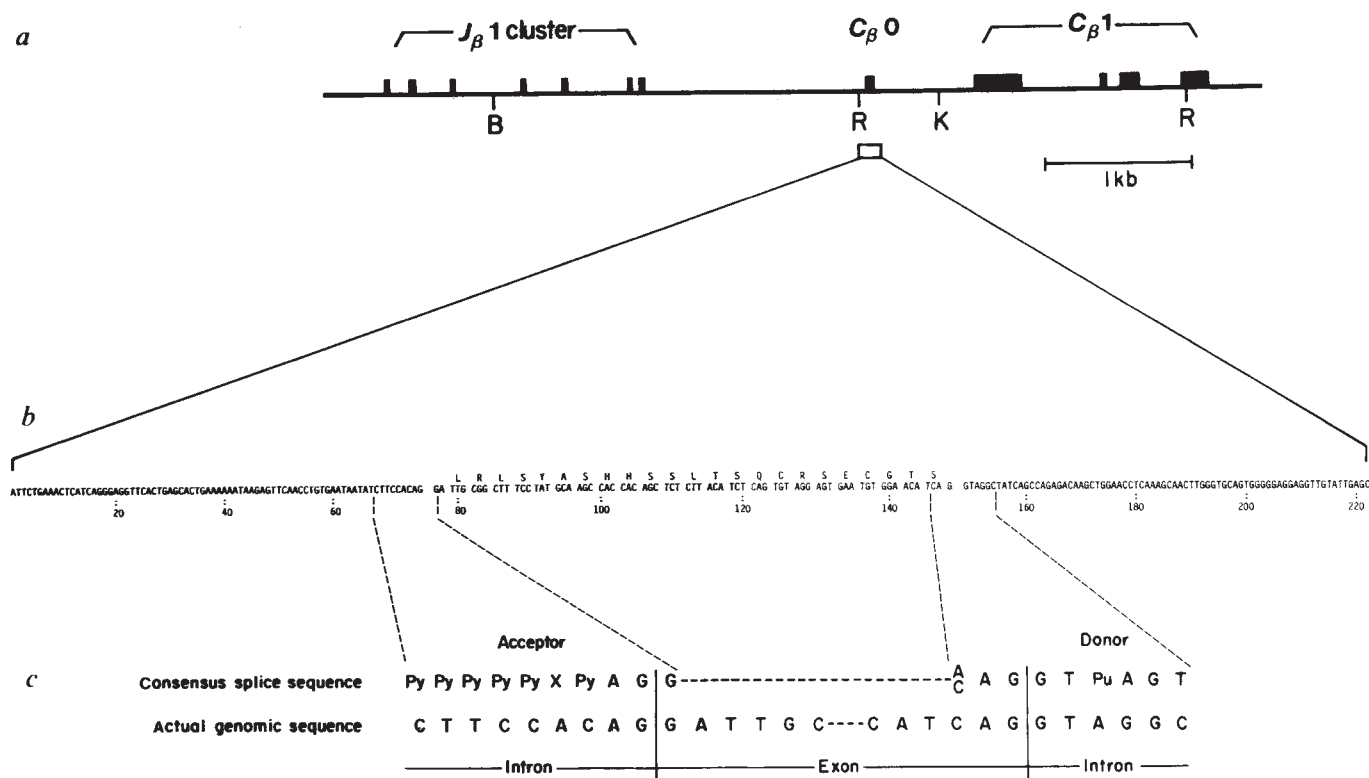


Fig. 2 Genomic localization of the $C_{\beta}0$ exon. **a**, Restriction map locating $C_{\beta}0$ between $J_{\beta}1$ and $C_{\beta}1$. B, BamHI; R, EcoRI; K, KpnI; kb, kilobase. **b**, Nucleotide sequence of the $C_{\beta}0$ exon and flanking introns. Nucleotide sequence is given beginning at the EcoRI site, as shown. **c**, Analysis of splice donor and acceptor sites. The nucleotide sequence of the $C_{\beta}0$ exon and its flanking introns is compared with the consensus donor and acceptor splice-site sequences reported by Breathnach and Chambon¹⁵. Py, pyrimidine; Pu, purine.

Methods. A BALB/c genomic library (EcoRI partial digest in Charon 4A, as described earlier²⁷) was screened with a C_{β} probe. Restriction maps of C_{β} -positive isolates were determined, and a clone was identified that was consistent with the map of the $C_{\beta}1$ locus reported earlier by Gascoigne *et al.*¹² Exon elements were located by hybridization. For **b**, the 2.2 kb EcoRI fragment containing $C_{\beta}0$ and $C_{\beta}1$ was subcloned into pUC12 and the nucleotide sequence of the indicated region was determined by the method of Maxam and Gilbert²⁵.

Expression of the $C_{\beta}0$ exon has been seen to vary from 1.3% of C_{β} -containing cDNA clones in a C57BL/6 spleen library to 18% of C_{β} -containing clones in an NZW thymus library. Direct comparison of the relative levels of $C_{\beta}0$ expression between these two libraries is difficult to interpret, however, because of uncertainty over the effects of the NZW $C_{\beta}1$ deletion on $C_{\beta}0$ expression. To better contrast the frequency of $C_{\beta}0$ expression between thymus and spleen, we constructed a C57BL/6 thymus cDNA library and screened a portion of the unamplified library with the C_{β} and $C_{\beta}0$ -splice oligonucleotide probes. Fourteen $C_{\beta}0$ -containing clones were identified from a total of 267 C_{β} -positives (Table 1). Thus 5% (14/267) of the C_{β} -containing clones in the C57BL/6 thymus library use the alternative splicing pathway.

$C_{\beta}0$ -containing cDNA clones represent a species of mRNA that is expressed at levels ranging from 5% to 13% of total C_{β} -containing clones in our C57BL/6 thymus and spleen libraries, respectively. A lower level of $C_{\beta}0$ expression in spleen than in thymus could be specifically regulated, or it could simply parallel a lower level of overall $C_{\beta}1$ usage in spleen. To address this possibility, the C57BL/6 thymus and spleen cDNA libraries were re-screened with probes specific for the 3'-untranslated sequence of $C_{\beta}1$ and $C_{\beta}2$; it was determined that $C_{\beta}1$ was present at levels of 32% and 27% of total C_{β} -positive clones in the thymus and spleen libraries, respectively (Table 1). The $C_{\beta}0$ element was present in 16% of thymic $C_{\beta}1$ clones and 5% of splenic $C_{\beta}1$ clones. No $C_{\beta}2$ clones were found to contain a $C_{\beta}0$ -hybridizing element. As the relative level of $C_{\beta}1$ usage does not change significantly between thymus and spleen, the decrease in $C_{\beta}0$ usage may result from some kind of specific

down-regulation of this pathway independent of $C_{\beta}1$ usage in general.

Interestingly, this is not the first report of an alternative RNA splicing pathway in a T-cell-specific transcript. The murine *Lyt 2* gene has been shown to produce two species of mRNA via alternative splicing²⁰. The relative proportions of the two mRNA species were seen to differ between thymus and spleen; in parallel with our observations of alternative splicing in TCR β -chain, expression of the minor species of *Lyt 2* mRNA was seen to be higher in thymus than in spleen.

Insertion of the $C_{\beta}0$ exon between a J_{β} element and the C_{β} element maintains the expected J_{β} - C_{β} translational reading frame, and the new exon itself encodes a 24-codon open reading frame. Thus, there is no obvious reason to suspect that $C_{\beta}0$ -containing β -chain transcripts should not be translated in parallel with other C_{β} -containing transcripts. While we have no data directly demonstrating the existence of a $C_{\beta}0$ -containing protein species, it seems reasonable to propose that such a species should actually exist, leading to the prediction that, in some circumstances, surface TCR could be a mixed population composed of both $C_{\beta}0$ -containing and $C_{\beta}0$ -excluding β -chain peptides.

Some predictions can be made from the primary sequence of the new exon regarding the potential effects its insertion may have on β -chain protein structure. The translated amino-acid sequences from Fig. 1 were analysed using the algorithm of Chou and Fasman²¹. The $C_{\beta}0$ sequence was found to occupy a region of low β -sheet and α -helix potentials with moderate reverse-turn probability, leading to a composite prediction of a random-turn secondary structure throughout this region. Inter-

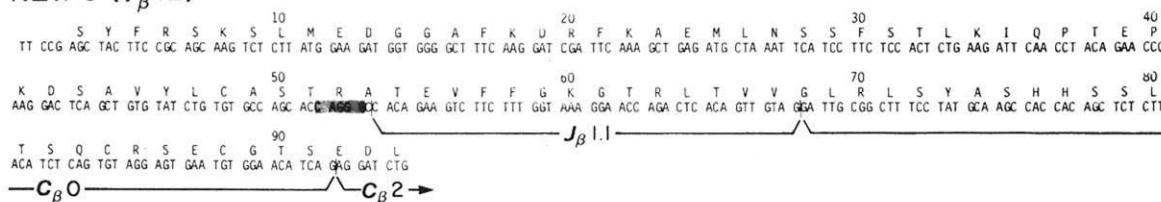
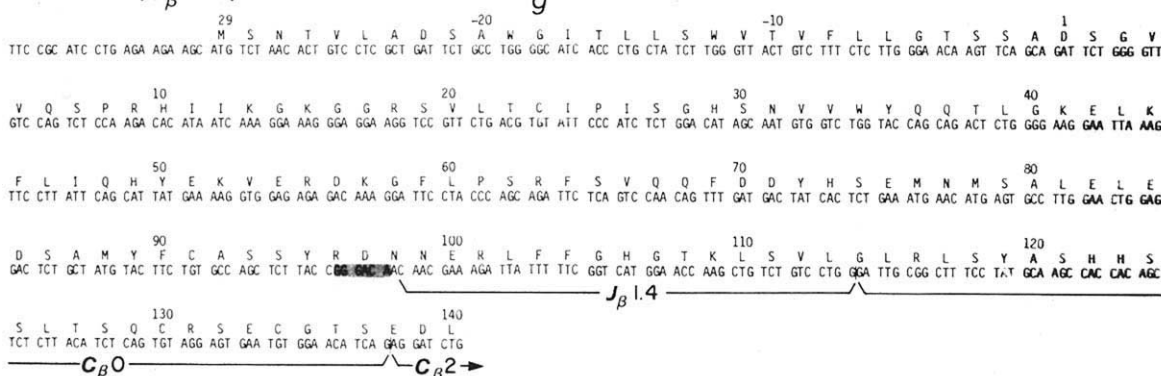
NZW 8 (V_{β} 12)NZW 22s (V_{β} 5.2)

Fig. 3 Nucleotide and amino-acid sequences of the V - D - J regions from two $C_{\beta}0$ -containing isolates from an NZW thymus cDNA library. Sequences contributed from germline $D_{\beta}1.1$ are shaded, while J_{β} , $C_{\beta}0$ and $C_{\beta}2$ gene segments are underlined. Nucleotides between D_{β} and J_{β} segments are proposed N-region insertions. The cDNA library was constructed from oligo(dT)-cellulose-selected NZW thymus RNA²⁶ in the λ gt10 cloning vector as described by Huynh *et al.*²⁸, with modifications²⁹. Library screening was done with ³²P-labelled oligonucleotide probes as described for Fig. 1c; $C_{\beta}0$ -positive clones were subcloned into pUC12 and their DNA sequences determined by the method of Maxam and Gilbert²⁵.

estingly, these parameters are very similar to those generated for the hinge region of human $\gamma 1$ immunoglobulins (data not shown). In addition, the $C_{\beta}0$ sequence contains two nearly adjacent cysteine residues in a region predicted to be hydrophilic by the algorithm of Hopp and Woods²². If these residues do not bond to each other, they may be available to form inter- or intra-chain disulphide bonds not usually present in the β -chain peptide. It is interesting to speculate that the presence of this sequence promotes the association of β -chain with a peptide other than α -chain (for example, γ -chain).

While it is not clear whether the $C_{\beta}0$ exon actually has any functional significance *in vivo*, it may be relevant to note that $C_{\beta}0$ is expressed at a fourfold higher level in thymus than in spleen. Decreased peripheral expression might simply reflect that few $C_{\beta}0$ -expressing T cells escape the thymus to become splenic T cells. Alternatively, it might reflect specific down-regulation of the $C_{\beta}0$ splicing pathway as cells leave the thymus. In regard to developmental regulation, we note that most $C_{\beta}0$ -containing cDNAs isolated from thymus are also V_{β} -containing,

suggesting that the more immature thymic ' D - J ' type transcripts²³ may express the $C_{\beta}0$ exon at lower levels than mature V - D - J transcripts. The thymus cDNA libraries examined here were constructed from tissue obtained from 6-week-old animals; it would be interesting to see whether the frequency of $C_{\beta}0$ expression varies with the developmental stage of the thymus. Further, we do not know whether the $C_{\beta}0$ splicing pathway is unique to the murine system or if it is also available to human TCR β -chain transcripts. In this regard, we note that no sequence with recognizable homology to $C_{\beta}0$ is present in the published human $C_{\beta}1$ sequence²⁴, although only a small portion of the intron between the $J_{\beta}1$ cluster and $C_{\beta}1$ has been reported.

We thank Mr Joseph Leykam for assistance with protein analysis and Madline Pearlman for manuscript preparation. This research was supported by funds from the Howard Hughes Medical Institute (D.Y.L.) and NIH Research Service Award GM 07200, Medical Scientist, from the National Institute of General Medical Sciences (M.A.B.).

Received 3 March; accepted 18 April 1986.

1. Takahashi, N., Roach, A., Teplow, D., Prusiner, S. & Hood, L. *Cell* **42**, 139-148 (1985).
2. deFerra, F. *et al. Cell* **43**, 721-727 (1985).
3. Amara, S., Jonas, V., Rosenfeld, M., Ong, E. & Evans, R. *Nature* **298**, 240-244 (1982).
4. Breitbart, R. *et al. Cell* **41**, 67-82 (1985).
5. Early, P. *et al. Cell* **20**, 313-319 (1980).
6. Moore, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 1800-1804 (1981).
7. Blattner, F. & Tucker, P. *Nature* **307**, 417-422 (1984).
8. Kotzin, B., Barr, V., Palmer, E. *Science* **229**, 167-171 (1985).
9. Noonan, D. *et al. J. exp. Med.* **163**, 644-653 (1986).
10. Behlke, M. *et al. Science* **229**, 566-570 (1985).
11. Chien, Y., Gascoigne, N., Kavalier, J., Lee, N. & Davis, M. *Nature* **309**, 322-326 (1984).
12. Gascoigne, N., Chien, Y., Becker, D., Kavalier, J. & Davis, M. *Nature* **310**, 387-391 (1984).
13. Barth, R. K. *et al. Nature* **316**, 517-523 (1985).
14. Behlke, M., Chou, H., Huppi, K. & Loh, D. *Proc. natn. Acad. Sci. U.S.A.* **83**, 767-771 (1986).
15. Breathnach, R. & Chambon, P. *A. Rev. Biochem.* **50**, 349-383 (1981).
16. Rupp, F., Acha-Orbea, H., Hengartner, H., Zinkernagel, R. & Joho, R. *Nature* **315**, 425-427 (1985).

17. Morinaga, T., Fortedat, A., Singh, B., Wegmann, T. & Tamaoki, T. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8163-8167 (1985).
18. Gorman, J. *et al. Cell* **40**, 859-867 (1985).
19. Malissen, M. *Cell* **37**, 1101-1110 (1984).
20. Zamoyska, R., Vollmer, A., Sizer, K., Liaw, C. & Parnes, J. *Cell* **43**, 153-163 (1985).
21. Chou, P. & Fasman, G. A. *Rev. Biochem.* **47**, 251-276 (1978).
22. Hopp, T. & Woods, K. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3824-3828 (1981).
23. Siu, G. *et al. Nature* **311**, 344-350 (1984).
24. Toyonaga, B., Yoshikai, Y., Vadasz, V., Chin, B. & Mak, T. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8624-8628 (1986).
25. Maxam, A. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-564 (1977).
26. Maniatis, T., Fritsch, E. & Sambrook, J. in *Molecular Cloning, A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
27. Loh, D., Bothwell, A., White-Scharf, M., Imanishi-Kari, T. & Baltimore, D. *Cell* **33**, 85-93 (1983).
28. Huynh, T., Young, R. & Davis, R. in *DNA Cloning: A Practical Approach* (ed. Glover, D.) 49-78 (IRL, Oxford, 1984).
29. Gubler, V. & Hoffman, B. *Gene* **25**, 263-269 (1983).
30. Roehm, N. *et al. J. Immun.* **135**, 2176-2182 (1985).
31. Staertz, U., Rammensee, H., Benedetto, J. & Bevan, M. *J. Immun.* **134**, 3994-4000 (1985).
32. Sim, G. & Augustin, A. *Cell* **42**, 89-92 (1985).

Synergism between immunoglobulin enhancers and promoters

J. Victor Garcia, Lê thi Bich-Thuy, Jeannine Stafford & Cary Queen

Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

Enhancers are DNA sequences that stimulate transcription from eukaryotic promoters¹⁻³. This stimulatory effect can be exerted over large distances and from a position either 5' or 3' of a promoter. Enhancers have been found in the genomes of many viruses, and in some cellular genes such as those encoding immunoglobulin heavy chain^{4,5} and κ light chain^{6,8}. An important feature of both viral and cellular enhancers is the ability of each enhancer to stimulate transcription from many promoters other than the one with which it is found associated^{1,2}. However, the question of whether cellular enhancers stimulate their 'own' promoter more efficiently than other promoters has apparently not been investigated. We show here that the κ light-chain enhancer stimulates a κ promoter about 20-fold more than it stimulates either the simian virus 40 (SV40) early promoter or a metallothionein (MT) promoter, two promoters that are very sensitive to other enhancers. Similarly, the heavy-chain enhancer stimulates a heavy-chain promoter much more than it stimulates the SV40 and MT promoters. This synergism between immunoglobulin enhancers and promoters might be due to the action of a protein that binds specifically to each of the regulatory elements.

To measure the effects of different enhancers, we used the chloramphenicol acetyltransferase (CAT) assay⁹, which has been widely employed for that purpose (see, for example, refs 10, 11). Four sets of plasmids were constructed. Plasmids used to determine the effect of enhancers on the SV40 early promoter were derived from pA10cat2 (ref. 10) which here is called pAcat (Fig. 1a). To determine the effect of enhancers on a κ gene promoter from the mouse myeloma MOPC 41 (ref. 12) we constructed the plasmid pKcat (Fig. 1b); in this plasmid, a fragment extending from an *EcoRI* site about 1,100 base pairs (bp) before the κ transcription start site to 25 bp beyond the start site was connected to the *cat* gene via a *BglII* adapter. Similarly, a fragment of the mouse 17.2.25 heavy-chain gene^{13,14} extending from ~2,000 bp before the promoter start site to 44 bp beyond the start site, was connected to the *cat* gene to form the plasmid pHcat. These plasmids also include the origin of replication of SV40 (ori), without the SV40 enhancer, in order to make them more analogous to pAcat (Fig. 1). Finally, the effect of enhancers on the mouse metallothionein-I (MT-I) promoter was determined using the plasmid pMcat. This plasmid is similar to pKcat and pHcat, except that the immunoglobulin gene segments are replaced by a fragment extending from an *EcoRI* site ~1,900 bp before the transcription start site of the MT gene to a *BglII* site located 64 bp after the start site¹⁵. On all plasmids, the initiating ATG of the *cat* gene is the first ATG codon after the expected transcription start site.

Three DNA fragments containing known enhancers from murine viruses or genes were inserted into each of the plasmids pAcat, pKcat, pHcat and pMcat at the sites indicated in Fig. 1. The polyomavirus and κ light-chain gene enhancers were, respectively, contained on the 238-bp and 473-bp fragments used previously¹⁶, and the heavy-chain enhancer was contained on a 682-bp *XbaI-EcoRI* fragment^{4,5}. The enhancers were inserted in the same orientation with respect to the direction of *cat* transcription that they have relative to the polyoma T-antigen, κ and heavy-chain genes. Plasmids containing the polyoma virus, κ and heavy-chain enhancers are designated with the suffixes P, K and H, respectively. All transfection experiments were performed on S194 cells (ATCC TIB 19), a mouse

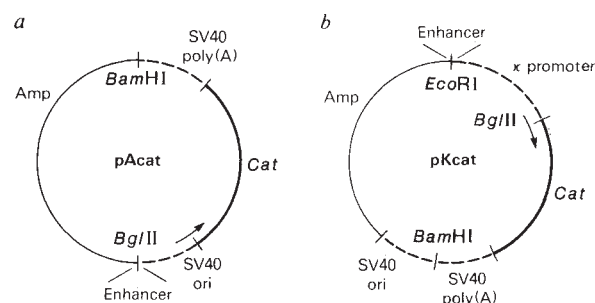


Fig. 1 Schematic diagrams of plasmids (not to scale). Arrows indicate the approximate initiation sites and direction of transcription from the SV40 early and κ promoters. Enhancers were inserted by means of *BglII* linkers into pAcat and by *EcoRI* linkers into pKcat. a, pAcat is the plasmid pA10cat2 of ref. 10, and does not contain the SV40 enhancer. b, The part of pKcat from the *EcoRI* site counterclockwise to the *BamHI* site was taken from pLT1 (ref. 6), and the part from the *BglII* site to the *BamHI* site is the *HindIII-BamHI* fragment of pSV2cat⁹. The plasmids pHcat and pMcat are similar to pKcat, with the κ gene segment replaced by heavy-chain and MT gene segments, respectively. Plasmids were constructed by standard methods; details available on request.

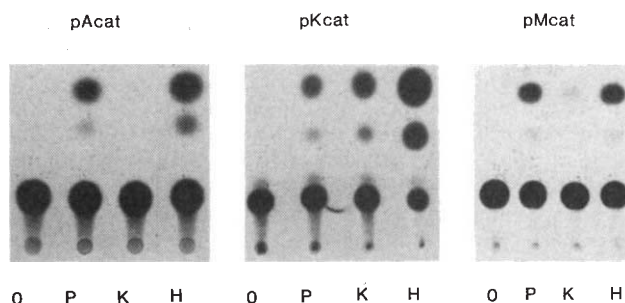


Fig. 2 CAT assays of extracts from transfected cells. The cells were transfected with the indicated plasmids containing the enhancers noted beneath the lanes: O (no enhancer), P (polyoma), K (κ) and H (heavy chain).

Methods. For each plasmid, one plate containing ~10⁷ S194 cells was transfected as described elsewhere⁸, except that the transfection was done in 1 ml with 5 μ g DNA. After 24 h, an additional 15 ml of medium was added to each plate, and CdCl₂ to 6 μ M was added to the cells transfected with the pMcat plasmids. Cells were harvested after another 24 h and CAT assays performed as described elsewhere⁹, using half of the extract from each plate and a 1-h incubation.

immunoglobulin (α , κ) myeloma line that we have found to have high transfection efficiency (unpublished data).

The plasmid pAcat and derivatives were transfected in parallel into S194 cells. After 48 h, the cells were lysed and tested for CAT activity⁹; Fig 2 shows the results of a typical experiment. The geometric means of the determinations of per cent ¹⁴C-chloramphenicol conversion for three separate experiments, using two sets of independently isolated plasmids, are shown in Table 1; all numbers have been normalized to the values for pAcatP rather than for pAcat, because the values for the latter are much smaller and may therefore be less reliable. Insertion of the κ enhancer into pAcat led to only a 4.2-fold increase (0.008 to 0.034) in transcription of *cat*, as measured by CAT synthesis, while the polyoma and heavy-chain enhancers respectively produced increases of 125- and 236-fold. These results are consistent with the finding that the κ enhancer is only 5% as effective as the heavy-chain enhancer in stimulating a β -globin promoter⁷. Two additional plasmids, identical to pAcatK and pAcatH except with the orientations of the enhancer fragments reversed, were transfected at the same time as the other plasmids,

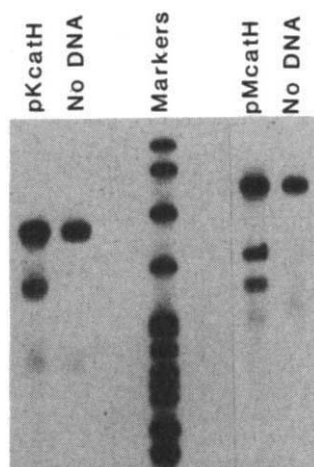


Fig. 3 S_1 -nuclease assay of RNAs from transfected cells. S194 cells were transfected with the indicated plasmids, containing the polyoma early region, or were mock-transfected without DNA. After 48 h, total RNA was extracted from the cells and assayed as described elsewhere⁸. The topmost band in each lane ran at the position of intact probe. The markers were an *Hpa*II digest of pBR322; sizes (in nucleotides) were: 622, 527, 404, 309, 242, 238, 217, 201, 190, 180, 160 and 147.

and yielded mean normalized CAT values of 0.038 and 0.70, respectively. Thus, reversing the orientation had no significant effect on the κ enhancer while it reduced the effect of the heavy-chain enhancer by a factor of ~ 3 .

Corresponding data were obtained for the plasmid pKcat, in which the κ promoter is linked to the *cat* gene (Fig. 2, Table 1). In contrast to the pAcat results, the κ enhancer stimulated transcription of the *cat* gene on pKcat more than did the polyoma enhancer, and the heavy-chain enhancer stimulated *cat* transcription by much more. In fact, the κ enhancer and the heavy-chain enhancer increased CAT synthesis by 88-fold (0.017 to 1.49) and 829-fold, respectively. Hence, the κ enhancer apparently stimulated transcription from the κ promoter about 20 times more efficiently than from the SV40 promoter, and the heavy-chain enhancer stimulated κ transcription about four times more than SV40 transcription. These results cannot be caused by the κ promoter being generally more sensitive to enhancers than the SV40 promoter, because the polyoma enhancer stimulated the κ promoter less than it did the SV40 promoter.

However, to verify that the phenomenon was not caused by a peculiarity of the SV40 promoter, we used pMcat, a plasmid very similar to pKcat in which a metallothionein promoter replaces the κ promoter. 24 h after transfection of pMcat and its derivatives into S194 cells, cadmium was added to the medium to induce the MT promoter. The results (Fig. 2, Table 1) were analogous to those for the pAcat plasmids. In fact, the κ en-

hancer stimulated transcription from the MT promoter only 4.6-fold, again about 20-fold less than from the κ promoter. Similarly, the heavy-chain enhancer increased transcription from the MT promoter much less than from the κ promoter. When cadmium was omitted from the medium, the amount of CAT synthesis from all the pMcat plasmids was decreased about fivefold, so that the amount of stimulation by the various enhancers remained approximately the same.

Finally, data were obtained from the pHcat plasmids, which have the heavy-chain promoter linked to *cat* (Table 1). The plasmid pHcat itself, which has no enhancer, expressed CAT at a very low level, precluding its accurate quantification. Moreover, whereas the heavy-chain enhancer stimulated the SV40 and MT promoters about as well as did the polyoma enhancer (Table 1), the heavy-chain enhancer stimulated the heavy-chain promoter 15.3 times more than did the polyoma enhancer. This result and the low level of heavy-chain promoter activity without any enhancer imply an especially effective interaction between the heavy-chain promoter and enhancer. Relative to the polyoma enhancer, the κ enhancer stimulated the heavy-chain promoter more (0.58) than it stimulated the SV40 or MT promoters (0.034 and 0.19) but less than it stimulated the κ promoter (1.49). Taking all the data together, the strongest synergism exists between the κ enhancer and promoter and between the heavy-chain enhancer and promoter, but there is also synergism between any pair of immunoglobulin gene elements.

Our results were not affected by fluctuations in transfection efficiency of the plasmids into the S194 myeloma cells. Indeed, each set of plasmids was independently transfected in parallel three times. The multiplicative standard deviations¹⁷ of the values for each plasmid were less than 1.6, except for pKcatH (1.82). In other experiments, the pKcat and pMcat plasmids were mixed with the plasmid pCH110, which contains a gene for β -galactosidase¹⁸, before transfection. Extracts of the transfected cells were assayed for both CAT and β -galactosidase activity. The β -galactosidase values for the four pKcat plasmids had a multiplicative standard deviation of only 1.02, and for the pMcat plasmids only 1.06. Hence, the transfection efficiencies for the plasmids within each experiment were very similar. Moreover, after normalization to the results for pKcatP and pMcatP, the CAT values for the plasmids carrying the κ and heavy-chain enhancers showed no significant differences from the values in Table 1.

Many studies have shown that enhancers do not alter the start point of transcription from the promoters that they stimulate (see, for example, refs 1, 4, 19). In particular, we have shown that in lymphoid cells, transcription initiates from the same place in the κ promoter when it is stimulated by either the κ or polyoma enhancers¹⁶. To further verify the location of the transcription start sites in the experiments presented here, we extracted RNA from S194 cells 48 h after they had been transfected with pKcatH or pMcatH. The RNAs were analysed using an S_1 -nuclease assay, with double-stranded probes labelled at the *Eco*RI site in the *cat* gene⁹, and respectively extending to the *Mae*III site 5' of the κ promoter²⁰ or to the *Sac*I site 5' of the MT promoter¹⁵. In initial S_1 experiments, no signals were detected clearly from pKcatH or pMcatH, probably because of instability of the *cat* RNA.

To increase the signal, we inserted the polyoma early region into pKcatH and pMcatH. This region has been shown to increase the amount of RNA transcribed from transfected plasmids by allowing them to replicate in mouse cells, thus increasing the amount of DNA template^{6,8,21}. The S_1 analysis of RNA from cells transfected with the modified pKcatH plasmid (Fig. 3) revealed a strong band of a length corresponding to transcripts from the usual κ start site⁶. The RNA from cells transfected with the modified pMcatH plasmid yielded two bands (in addition to residual intact probe): the length of the stronger, upper band corresponds to transcripts from the usual MT start site¹⁵;

Table 1 Effect of enhancers on different promoters

Enhancer	Plasmid			
	pAcat	pKcat	pMcat	pHcat
None	0.008	0.017	0.041	<0.01
Polyoma	1.00	1.00	1.00	1.00
κ	0.034	1.49	0.19	0.58
Heavy chain	1.89	14.1	0.92	15.3

Each value is the geometric mean of determinations of per cent chloramphenicol conversion after three independent transfections, and is normalized to the values for the polyoma enhancer.

the lower band presumably represents transcripts initiating ~40 bp farther downstream. Both bands were also present when the MT promoter was stimulated by the polyoma enhancer in pMcatP, but were much weaker when there was no enhancer present on the plasmid (not shown). The shorter transcript from the MT promoter has not been observed previously¹⁵, but the probe used in that study would probably not have detected it.

An attractive explanation for the observed synergism between immunoglobulin enhancers and promoters is that one or more proteins specifically bind to all of these regulatory regions. The existence of such proteins might be reflected in sequence homologies between the regulatory regions. Such homologies

have indeed been observed²². Recently, direct evidence of a protein that binds both immunoglobulin promoters and enhancers has been obtained²³. The results presented here may help to explain the negative findings of searches for cellular enhancers that use a heterologous promoter as the assay²⁴. Many enhancers in cellular genes may, like the κ enhancer, efficiently stimulate only their own or related promoters.

We thank L. Laimins, F. Lee and D. Hamer for providing pAcat, pCH110 and the MT promoter, respectively. We thank C. Mock and B. Miller for photographic and secretarial assistance, and M. Singer for reviewing the manuscript. L.T.B.-T. was partially supported by INSERM, France.

Received 16 April; accepted 20 May 1986.

1. Banerji, J., Rusconi, S. & Schaffner, W. *Cell* **27**, 299–308 (1981).
2. Moreau, P. *et al. Nucleic Acids Res.* **9**, 6047–6068 (1981).
3. Fromm, M. & Berg, P. *J. molec. appl. Genet.* **1**, 457–481 (1982).
4. Banerji, J., Olson, L. & Schaffner, W. *Cell* **33**, 729–740 (1983).
5. Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. *Cell* **33**, 717–728 (1983).
6. Queen, C. & Baltimore, D. *Cell* **33**, 741–748 (1983).
7. Picard, D. & Shaffner, W. *Nature* **307**, 80–83 (1984).
8. Queen, C. & Stafford, J. *Molec. cell. Biol.* **4**, 1042–1049 (1984).
9. Gorman, C., Moffat, L. & Howard, B. *Molec. cell. Biol.* **2**, 1044–1051 (1982).
10. Laimins, L. A., Gruss, P., Pozzatti, R. & Khoury, G. *J. Virol.* **49**, 183–189 (1984).
11. Herbolmel, P., Bourachot, B. & Yaniv, M. *Cell* **39**, 653–666 (1984).
12. Seidman, J. G. & Leder, P. *Nature* **276**, 790–795 (1978).

13. Loh, D. Y., Bothwell, A. L. M., White-Scarff, M. E., Imanishi-Kari, T. & Baltimore, D. *Cell* **33**, 85–93 (1983).
14. Grosschedl, R. & Baltimore, D. *Cell* **41**, 885–897 (1985).
15. Hamer, D. H. & Walling, M. *J. molec. appl. Genet.* **1**, 273–288 (1982).
16. Foster, J., Stafford, J. & Queen, C. *Nature* **315**, 423–425 (1985).
17. Snedecor, G. W. & Cochran, W. G. *Statistical Methods* 6th edn, 330 (Iowa State University Press, 1967).
18. Hall, C., Jacob, E., Ringold, G. & Lee, F. *J. molec. appl. Genet.* **2**, 101–109 (1983).
19. Levinson, B., Khoury, G., Vande Woude, G. & Gruss, P. *Nature* **295**, 568–572 (1982).
20. Seidman, J. G., Max, E. E. & Leder, P. *Nature* **280**, 370–375 (1979).
21. Deans, R., Denis, K. A., Taylor, A. & Wall, R. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1292–1296 (1984).
22. Falkner, F. G. & Zachau, H. G. *Nature* **310**, 71–74 (1984).
23. Singh, H., Sen, R., Baltimore, D. & Sharp, P. A. *Nature* **319**, 154–158 (1986).
24. Weber, F., de Villiers, J. & Schaffner, W. *Cell* **36**, 983–992 (1984).

Expression of human adenosine deaminase in murine haematopoietic progenitor cells following retroviral transfer

John W. Belmont*, Jenny Henkel-Tigges*, Stephen M. W. Chang*, Karen Wager-Smith*, Rodney E. Kellems†, John E. Dick‡, M. Cristina Magli‡, Robert A. Phillips‡, Alan Bernstein‡ & C. Thomas Caskey*

* Institute for Molecular Genetics and Howard Hughes Medical Institute and † Department of Biochemistry, Baylor College of Medicine, Houston, Texas 77030, USA

‡ Ontario Cancer Institute and Mt Sinai Hospital Research Institute, Department of Medical Genetics and Biophysics, University of Toronto, Toronto, Ontario, Canada M4X 1K9

Adenosine deaminase (ADA) deficiency, an autosomal recessive inborn error of metabolism, leads to severe combined immune deficiency in man¹. This enzyme, although constitutively expressed in most tissues, is expressed at high level in immature T cells, and study of the pathophysiology of the disorder indicates that increased deoxyadenosine or altered methylation capacity have toxic effects on T-cell maturation¹. Although bone marrow transplantation can correct the immune deficiency^{2,3}, this therapy is associated with graft-versus-host disease and incomplete immune restoration, and so our laboratory and others have sought to develop a method of gene replacement as a possible treatment for the disease⁴. Moreover, characterization of the complementary DNA of the human ADA gene and some of its mutants^{5,6} makes it possible to design gene transfer strategies. We have now subcloned a human adenosine deaminase cDNA into the retrovirus shuttle vector pZIP-SV(B), and in this way have isolated a cell line, 4.2T, which produces high titres of replication-defective retrovirus which have been used to transfer the gene for human ADA to mouse bone marrow cells. Transfer and expression of the neomycin-resistance gene (*neo*) and the ADA gene in murine bone marrow colony-forming units (CFU) was demonstrated by *in vitro* colony formation in the presence of the antibiotic G418 or 9-xylofuranosyladenine plus deoxycytosine, respectively. Isoenzyme analysis also showed human ADA expression in the cultured mouse bone marrow.

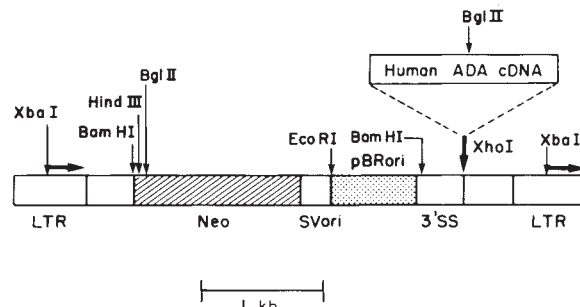


Fig. 1 Construction of the 4.2T cell line. The SV(B) plasmid vector was cleaved with *Xho*I, filled in with Klenow and treated with calf intestinal phosphatase (CIP). The pADA211⁵ plasmid containing a full-length human ADA cDNA was cleaved with *Eco*RI and *Dde*I, thus removing the 3' poly(A) addition site, filled in with Klenow and blunt-end ligated to the SV(B) vector. The SVBADA211 plasmid was isolated by banding in CsCl, and 10 μ g purified DNA was used to transform PA-12 (ref. 9) cells by CaPO₄ precipitation. Supernatant from these cells, containing the transiently expressed SVBADA211 virus with an amphotropic envelope, was collected at 48 h and used to infect ψ_2 cells. These cells were placed in Dulbecco's modified Eagle's medium (DMEM) with 10% FCS containing G418 (300 μ g ml⁻¹). Ninety-six G418-resistant clones were selected and placed in 11AAU (1.1 mM adenosine, 50 mM alanosine, 1.0 mM uridine) plus 50 nM deoxycytosine (dCf). Two cell lines, 4.1T and 4.2T, were able to grow in this selective medium, and 4.2T had both higher expression of human ADA and higher titre virus production. LTR, long terminal repeat. SVori, SV40 origin of replication; pBRori, origin of replication from pBR322; 3' SS, 3' splice site; kb, kilobases.

The development of a biological gene transfer system based on highly infectious retroviruses opens the way to efficiently introducing genes into rare cell populations^{7–10}, and several studies have recently demonstrated that the *neo* gene can be transferred into murine haematopoietic progenitors and stem cells using retroviral vectors^{11–15}. However, ADA gene transfer is hampered by the need to fulfil several requirements: (1) the production of a high-titre ADA vector without concomitant production of replication-competent virus, (2) expression of the transferred ADA gene in the target cells (that is, in marrow stem cells and their differentiated progeny), and (3) adequate expression at the tissue level in the test animal (ADA in thymus

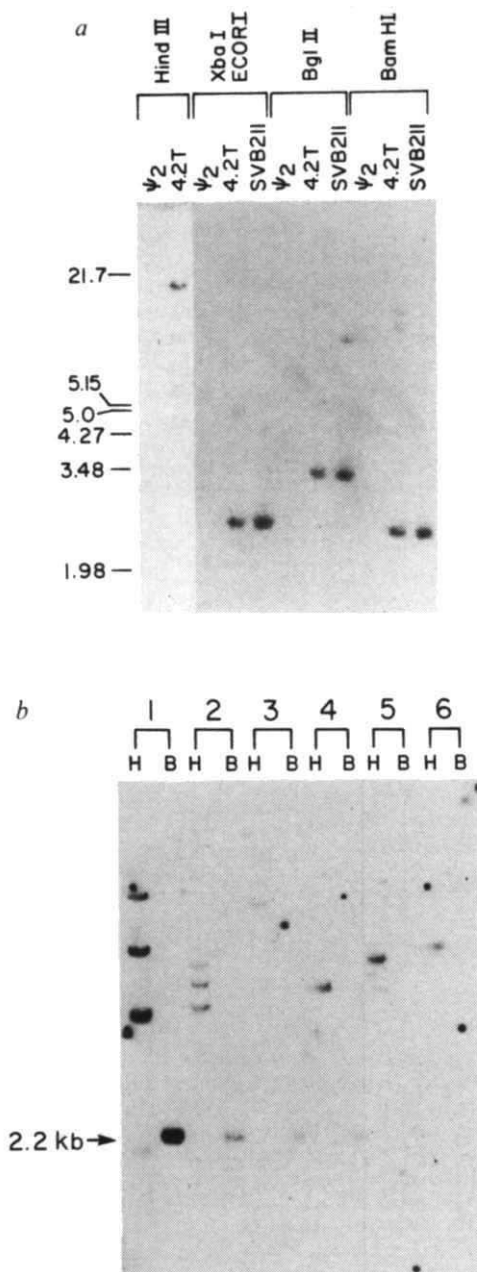


Fig. 2 *a*, Southern analysis of 4.2T. *b*, Southern analysis of 4.2T-infected cells. H, digestion with *Hind*III; B, digestion with *Bam* I. Lanes 1-6 represent individual clonal cell lines.

Methods. *a*, Genomic DNA was prepared from cultured ψ_2 and 4.2T cells by standard techniques. 10 μ g of genomic DNA or 40 pg of plasmid was digested overnight with the indicated restriction enzymes, separated on 0.8% agarose, Tris-acetate-EDTA buffered gel at 40 V for 16 h. The gels were blotted onto nylon membrane, prehybridized overnight and probed with nick-translated *neo* insert at 42°C with 50% formamide. The filters were washed in 0.1 $\times$ SSC at 60°C and then autoradiographed overnight. *b*, NIH 3T3 cells were infected with supernatants from 4.2T. After 48 h the cells were placed in DMEM with 300 μ g ml⁻¹ G418. The surviving cells were clonally selected and grown in DMEM plus G418. Genomic DNA was extracted, digested with either *Hind*III (H) or *Bam*HI (B). Filter hybridization with ³²P-*neo* was as described in A. Cell lines with apparent multiple inserts (lanes 1, 2) are being recloned to establish homogeneity.

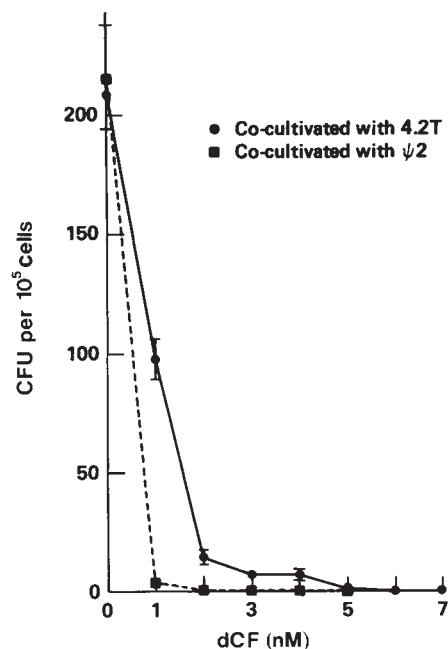


Fig. 3 Selection of CFU for ADA expression. Co-cultivation of 4.2T and mouse marrow cells was identical with that described in Table 1. After 24 h co-cultivation, the non-adherent cells were collected and plated in IMDM with 1% methylcellulose cultures with 5 μ M XylA and varying concentrations of dCF. CFU were counted at 7 days. Data represent averages of duplicate culture from two separate experiments $\pm$ s.e.m.

and spleen). The present experiments investigated these problems in committed bone marrow progenitor cells. The vector pZIP-SV(B)¹⁰ (Fig. 1) was chosen because it allows the *neo* gene to be used as a dominant selectable marker for expression. The cell line 4.2T was formed by infection of the ψ_2 packaging line by defective amphotropic SVBADA211 virus. This line was clonally isolated after sequential selection in G418¹⁶ and then 11AAU (1.1 mM adenosine, 50 mM alanosine, 1.0 mM uridine) plus 50 nM deoxycoformycin (dCF)¹⁷. These selections require expression of neomycin phosphotransferase and increased ADA expression. The amount of dCF added makes the selection very stringent and thus rapidly identifies cells showing maximal expression of the transferred ADA gene. Of 96 G418-resistant isolates screened in this way, the 4.2T line had the highest expression of human ADA. The titre on NIH 3T3 mouse fibroblasts is $\sim 2 \times 10^4$ CFU per ml (G418 resistance). Southern analysis of 4.2T (Fig. 2a) indicated that the cell line has a single proviral insert. In addition, the 4.2T provirus was identical with plasmid pSVBADA211 using several internal restriction cuts, indicating that the selection procedure did not engender a gross rearrangement of the provirus. Southern analysis of cell lines infected with virus produced by 4.2T showed intact internal provirus structure (Fig. 2b). This also strongly suggests that the basic transcriptional unit of the provirus was not altered during selection. A more subtle alteration—for example, point mutations—cannot be ruled out, however.

The 4.2T cell line was then used to test the transfer and expression of the *neo* and ADA genes in murine haematopoietic cells. Mouse bone marrow was co-cultivated with 4.2T cells for 24 h according to the protocol of Dick *et al.*¹³. Non-adherent cells were replated in liquid media for 48 h with or without preselection in G418. These cells were then replated in semisolid media in the absence or presence of G418 to determine the proportion of resistant CFU. Table 1 indicates that approximately 10% of CFU acquire G418 resistance when co-cultivated with 4.2T. Co-cultivation with ψ_2 gives no G418-resistant colonies, indicating the specificity of the effect. Metabolic selec-

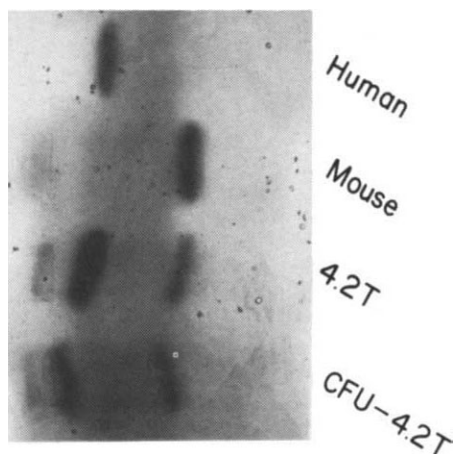


Fig. 4 Isoenzyme analysis of ADA in G418-selected infected bone marrow cells. Lane 1, HeLa cell (human) extract; lane 2, ψ_2 cell (mouse) extract; lane 3, 4.2T; lane 4, CFU co-cultivated with 4.2T cells. The infection of mouse bone marrow cells was identical with that described in Table 1 except that the 4.2T cells were irradiated (1,500R, Gamma-cell 100, ^{137}Cs source) before co-cultivation. Cells from CFU cultures were collected by repeated washing to remove the methylcellulose. Cell extracts were made by freezing and thawing in 20 mM Tris-HCl, pH 7.5. Supernatants were cleared by 15 min spin in a microcentrifuge. Cell extracts were loaded on a prefocused 1-mm-thick pH 4–5 polyacrylamide isoelectric focusing gel. Focusing was stopped after 3–4 h and the gel was soaked in 50 mM Tris-HCl pH 8.0 for 20 min. The gel was then stained for ADA enzyme activity using an agar overlay by standard methods¹⁷.

tion of CFU for increased expression of ADA was also performed using 9-xylofuranosyl adenine (XylA) plus dCF¹⁸ in a similar protocol (Fig. 3). XylA plus dCF was used for ADA selection because it gives a sharper inflection on survival curves than 11AAU for both fibroblasts and CFU. These data show a preferential survival of CFU co-cultivated with 4.2T compared with those co-cultivated with ψ_2 . At medium concentrations of dCF (2–4 nM), the per cent resistant CFU was comparable to the results obtained with G418 selection. To confirm the expression of human ADA in these cultured marrow cells, the cells were pooled from G418-selected cultures, and freeze-thaw extracts were prepared and run on an isoelectric focusing gel (Fig. 4). A band corresponding to the human ADA was detected only in CFU arising from 4.2T-infected marrow. Carryover of human enzyme activity from 4.2T virus-producer cells is highly unlikely since the monolayer was irradiated and there were two rounds of removal of adherent cells before replating in semisolid media. However, we cannot rule out a contribution to the observed expression by more mature non-colony-forming marrow cells.

These preliminary studies show that by exploiting the selectability of the ADA gene, cell lines producing high-titre retrovirus can be rapidly selected. The endogenous ADA gene can be amplified using 11AAU plus dCF¹⁷, and it may therefore be possible to increase virus production by amplification of the ADA-containing provirus. We have also shown that this particular SV(B)-human ADA construction is expressed well in murine haematopoietic cells after infection *in vitro*. There are several reports indicating problems with expression in retroviral constructs containing two genes^{19,20}. Furthermore, expression of a particular retroviral construct in mature marrow cells and committed progenitors does not guarantee that the same retroviral vector will function well after infection of more primitive stem cells. Indeed, preliminary data indicate significant loss of expression of *neo* in the differentiated progeny of infected CFU (ref. 21 and J.E.D. *et al.*, manuscript in preparation). It is generally considered that constructs that use separate promoters

Table 1 Selection of CFU for *neo* expression

Preselection	Selection	ψ_2	4.2T
—	—	240 ± 26	218 ± 20
—	G418	0	24 ± 5
G418	—	0	19 ± 5
G418	G418	0	7 ± 2

ψ_2 or 4.2T cells (1×10^6) were plated in 100-mm tissue culture dishes (Corning) in Iscove's modified Dulbecco's medium (IMDM, Gibco) with 25% fetal calf serum (FCS). The next day 4×10^6 bone marrow cells from the tibias and femurs of C3H/HeJ or C57B16/J (Jackson Laboratories) were placed in each dish in IMDM supplemented with bovine serum albumin (BSA), soybean lipid, Fe-saturated transferrin (Boehringer Mannheim), penicillin-streptomycin (Gibco), 10% FCS, 10% WEHI3B-conditioned medium, α -thioglycerol and $2 \mu\text{g ml}^{-1}$ Polybrene (Sigma). Twenty-four hours later, non-adherent cells were washed off and plated in fresh 100-mm Petri dishes (Falcon) with or without selective drugs (preselection). After 48 h recovery, the non-adherent cells were again collected, counted and 1×10^5 cells were plated in 1 ml IMDM with 1% methylcellulose (Fluka). These cultures had either no drug selection, or 1 mg ml^{-1} G418. CFU (>200 cells) were counted 7–10 days after initiation of the semisolid cultures. Data are expressed as averages of three separate experiments $\pm$ s.e.m.

for each cDNA (two transcriptional units) give better expression in target cells. The SV(B) construction used here has only one transcriptional unit. We presume that expression in the target cell is dependent on the particular features of each construction and that it is premature to predict how a construction will behave at the stages of virus production and target cell expression. Other constructs using both double-promoter and unspliced designs are under investigation in our laboratory.

Somatic gene transfer techniques may allow a number of difficult biological problems to be addressed. Several reports indicate its usefulness in investigating cell lineage relationships in the haematopoietic system^{13,14}, and similar hierarchy studies may be undertaken with other tissues if infection is carried out at multiple developmental stages. Retrovirus gene transfer may also be useful for the study of oncogenesis^{22,23}, while transfer of poorly understood genes into haematopoietic cells (for example, T1a or T11 antigen) may help to clarify their functional role. Finally, this technique may allow safe and effective treatment of ADA deficiency and various other inborn errors of metabolism by somatic gene therapy.

Received 2 May; accepted 5 June 1986.

- Kredich, N. & Herschfield, M. *The Metabolic Basis of Inherited Disease*, 1157–1183 (McGraw-Hill, New York, 1983).
- Parkman, R., Gelfand, E. W., Rosen, F. S., Sanderson, A. & Hirshhorn, R. *New Engl. J. Med.* **292**, 714–719 (1975).
- Polmar, S. H. *et al.* *New Engl. J. Med.* **295**, 1337–1343 (1976).
- Anderson, W. F. *Science* **226**, 401–409 (1984).
- Adrian, G. S., Wiginton, D. A. & Hutton, J. J. *Molec. cell. Biol.* **4**, 1712–1717 (1984).
- Bonthron, D. T., Macham, A. F., Ginsburg, D. & Orkin, S. H. *J. clin. Invest.* **76**, 894–897 (1985).
- Bernstein, A., Berger, S., Huszar, D. & Dick, J. *Genetic Engineering, Principles and Methods* Vol. 7, 235–261 (Plenum, New York, 1985).
- Mann, R., Mulligan, R. C. & Baltimore, D. *Cell* **33**, 153–159 (1983).
- Miller, A. D., Low, M.-F. & Verma, I. M. *Molec. cell. Biol.* **5**, 431–437 (1985).
- Cepko, C. L., Roberts, B. E. & Mulligan, R. C. *Cell* **37**, 1053–1062 (1984).
- Joyner, A., Keller, G., Phillips, R. A. & Bernstein, A. *Nature* **305**, 556–558 (1983).
- Williams, D. A., Lemischka, I. T., Nathan, D. G. & Mulligan, R. C. *Nature* **310**, 476–480 (1984).
- Dick, J. E., Magli, M. C., Huszar, D., Phillips, R. A. & Bernstein, A. *Cell* **42**, 71–79 (1985).
- Keller, G., Paige, C., Gilboa, E. & Wagner, E. *Nature* **318**, 149–154 (1985).
- Eglitis, M. A., Kantoff, P., Gilboa, E. & Anderson, W. F. *Science* **230**, 1395–1398 (1985).
- Colbère-Garapin, F., Horodniceanu, F., Kourilsky, P. & Garapin, A. C. *J. molec. Biol.* **150**, 1–14 (1981).
- Yeung, C.-Y. *et al.* *J. biol. Chem.* **258**, 8338–8345 (1983).
- Yeung, C.-Y., Riser, M., Kellems, R. E. & Siciliano, M. J. *Molec. cell. Biol.* **3**, 2191–2202 (1983).
- Emerman, M. & Temin, H. M. *Cell* **39**, 459–467 (1984).
- Joyner, A. L. & Bernstein, A. *Molec. cell. Biol.* **3**, 2191–2202 (1983).
- Williams, D. A., Orkin, S. H. & Mulligan, R. C. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2566–2570 (1986).
- Wolf, L. & Ruscetti, S. *Science* **228**, 1549–1552 (1985).
- Berger, S. A., Sanderson, N., Bernstein, A. & Hawkins, W. D. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6913–6917 (1985).

The alternative nitrogenase of *Azotobacter chroococcum* is a vanadium enzyme

Robert L. Robson, Robert R. Eady, Toby H. Richardson,
Richard W. Miller*, Marie Hawkins
& John R. Postgate

AFRC Unit of Nitrogen Fixation, University of Sussex,
Brighton BN1 9RQ, UK

*Chemistry and Biology Research Institute, Agriculture Canada,
Ottawa, Canada K1A 0C6

The requirement for molybdenum in biological dinitrogen fixation, first reported by Bortels¹, is due to its involvement at or near the site of reduction of N_2 in conventional nitrogenase. To date, all nitrogenases which have been purified to homogeneity consist of an iron protein (component 2) and a molybdoprotein (component 1)². *Azotobacter vinelandii*, an obligately aerobic diazotrophic bacterium, has two systems for nitrogen fixation: a conventional nitrogenase involving molybdenum and an alternative system which functions under conditions of Mo deficiency and does not require the structural genes for conventional nitrogenase³⁻⁶. The properties of the nitrogenase in extracts of comparable deletion strains of *A. vinelandii* are consistent with a two-component system^{6,7} in which the component 1-containing fraction has no detectable Mo (ref. 6). Recently, an alternative nitrogen fixation system has been demonstrated in *Azotobacter chroococcum* strain MCD1155, in which the structural genes for conventional nitrogenase are deleted⁸. We demonstrate here that nitrogen fixation by this strain depends on vanadium and we show that its purified nitrogenase is a binary system in which the conventional molybdoprotein is replaced by a vanadoprotein.

A. chroococcum NCIMB 8003 contains conventional nitrogenase when grown diazotrophically in a medium containing molybdate⁹. The genes which encode the polypeptides for this nitrogenase (*nifH* for the Fe-protein, Ac2; *nifD* and *nifK* for the α and β subunits of the MoFe-protein, Ac1) have been cloned as a contiguous cluster from this organism¹⁰. Strains in which this gene cluster was deleted, constructed by a gene-replacement technique, were incapable of N_2 fixation when molybdate was provided, but developed feeble acetylene reduction activity indicative of nitrogenase in Mo-deficient medium⁸. A derivative, MCD1155, defective in uptake hydrogenase activity and resistant to up to 5 mM sodium tungstate (an inhibitor of molybdate uptake or assimilation in azotobacter^{11,12}), grew in N_2 , and exhibited ^{15}N incorporation from $^{15}N_2$, acetylene reduction and nitrogenase-dependent H_2 evolution. Growth and nitrogenase activity in MCD1155 apparently required relatively high (2.5 mM) concentrations of sodium tungstate⁸.

The most probable reason for so high a tungstate requirement is that our tungstate was contaminated with traces of an element (or elements) required for diazotrophy in this strain. We therefore tested cultures which grew poorly, in the N-deficient, Mo-deficient medium of Fig. 1, for stimulation of growth by Na_2SeO_3 (29 μM), $MnCl_2$ (15.2 μM), $CuSO_4$ (0.6 μM), $NiCl_2$ (0.2 μM), $CoCl_2$ (5 μM), $FeSO_4$ (50 μM), $ZnSO_4$ (7 μM), $CaCl_2$ (100 μM), Na_2MoO_4 (10 nM or 100 μM), Na_3BO_3 (81 μM) or $NaReO_4$ (10 μM). None stimulated growth. However, $VOSO_4$ at low concentrations restored growth; Fig. 1a shows that nitrogenase activity as measured by H_2 evolution was proportional to added $VOSO_4$ between 1 and 10 nM. From these data we deduced a specific activity for H_2 evolution in air of 120 pmol per min per pg atoms V. Mo, V-starved organisms responded rapidly and in a biphasic pattern to V (Fig. 1b); a rapid two fold activation of hydrogen-evolving activity seen after 30 min was followed by a slower progressive increase in activity over 10 h. Growth was evident after 6 h. We conclude that vanadium is required for diazotrophic growth of this strain.

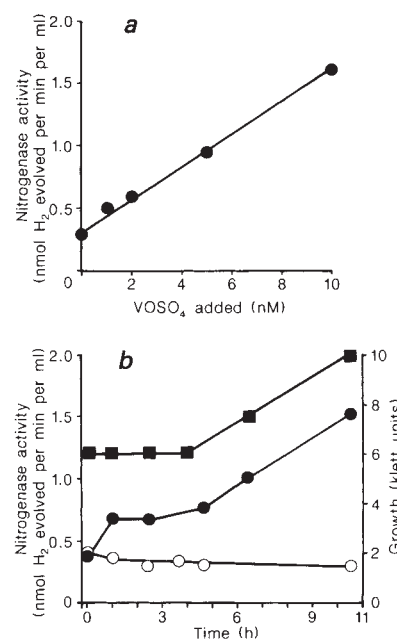


Fig. 1 Influence of vanadium on diazotrophic growth and nitrogenase activity in *A. chroococcum* MCD1155. *a*, Organisms were inoculated to 5×10^6 cells ml^{-1} . $VOSO_4$ was added and nitrogenase activities were measured in cultures by following H_2 evolution. Maximum activities obtained after 48 h at 30 °C are plotted. *b*, Organisms were inoculated to 5×10^7 ml^{-1} . At time zero, $VOSO_4$ was added to 10 nM. Nitrogenase activity was determined at intervals by following H_2 evolution over 15 min. Growth was determined using a Klett-Summerson photoelectric colorimeter and is shown for the V-supplemented culture only (■). No growth was observed without added V. ●, Nitrogenase activity with V added; ○, without V added.

Methods. *A. chroococcum* MCD1155 was routinely grown in air at 30 °C in (g per l of distilled H_2O): sucrose 20; $K_2HPO_4 \cdot 3H_2O$, 0.64; KH_2PO_4 , 0.16; Na_2SO_4 , 0.142; $MgCl_2 \cdot 6H_2O$, 0.203; $CaCl_2 \cdot 2H_2O$, 0.074; $FeC_6H_5O_7$, 0.0335; Na_2WO_4 , 0.734. To test for the influence of V on growth and nitrogenase activity, organisms were grown without added tungstate until residual level was 2.5 μM .

The apparent requirements for high levels of tungstate could have been due to: (1) competitive exclusion of traces of molybdenum, (2) use of W in place of V via an inefficient uptake system or (3) contamination of our tungstate source with traces of V. We eliminate possibility (1) because, in shaken flasks with the Mo-free medium, diazotrophic growth of MCD1155 depended on tungstate concentration up to 2.5 mM, but addition of normal levels of molybdate (100 μM) to cultures containing 2.5 mM tungstate did not inhibit growth. Our studies do not conclusively eliminate explanation (2), but we favour explanation (3) because we have detected trace contamination (22 p.p.m. V) of our tungstate source and, in the presence of 2.5 mM tungstate, added vanadium did not stimulate growth.

Nitrogenase activity in extracts of MCD1155 prepared as in Fig. 2 had properties and requirements similar to those of conventional nitrogenases. Activity measurable as H_2 evolution, acetylene reduction or ammonia production was ATP- and reductant-dependent (data not shown). The activity was highly sensitive to irreversible inactivation by O_2 , unlike the enzyme in crude extracts of the parent strain grown in molybdenum-sufficient conditions.

MCD1155 nitrogenase was purified essentially as described in Fig. 2 legend. Two readily separable yellow-brown coloured protein components were purified to homogeneity by a combination of ion-exchange and gel-filtration chromatography (Fig. 2). Both proteins were essential for activity. We refer to these

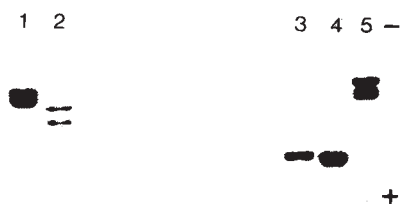


Fig. 2 Polyacrylamide gel electrophoresis of nitrogenase components of *A. chroococcum* MCD1155. Lanes 1 and 5, Ac1 2.9 µg and 2.9 µg, respectively; lane 2, Ac1* 2 µg; lane 3, Ac2* 11.0 µg; lane 4, Ac2 2.7 µg.

Methods. Organisms were grown in 400 l of the medium described in Fig. 1 legend with 40 µM VOSO₄ added. Cells were collected at mid-exponential growth, resuspended in 50 mM HEPES buffer pH 8.0 and disrupted by French pressure cell treatment. Nitrogenase was purified under strictly anoxic conditions and separated into two components using a combination of DEAE-Sephacel chromatography in 50 mM Tris-HCl buffer pH 8.0 and gel filtration in 50 mM Tris-HCl buffer pH 8.0, 50 mM MgCl₂ essentially as described for nitrogenase from *Klebsiella pneumoniae*²⁶. Both proteins were then subjected to a linear NaCl gradient (0.25–0.4 M NaCl; total volume 300 ml) on DEAE-Sephacel equilibrated with 50 mM Tris-HCl pH 8.0 to remove contaminating proteins. Purified components were subjected to treatment with 1% SDS and 1% β-mercaptoethanol, electrophoresed as described previously²⁷ and stained with Coomassie brilliant blue. Nitrogenase components (Ac1* and Ac2*) are compared with samples of conventional nitrogenase (Ac1 and Ac2) purified from *A. chroococcum* grown in Mo-sufficient conditions⁹.

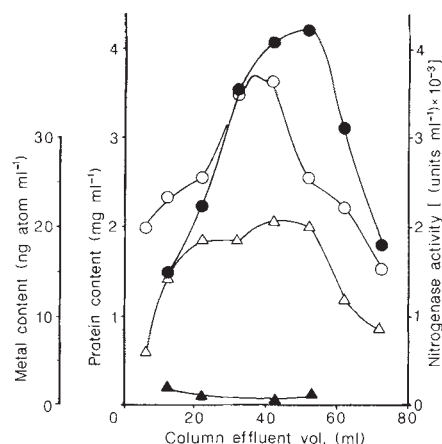


Fig. 3 Correlation of vanadium content of Ac1* with nitrogenase activity. The figure shows the elution profile of the larger component of nitrogenase (Ac1*, M_r 210,000) from DEAE-Sephacel under the conditions described in Fig. 2 legend. Nitrogenase activity (●), Mo (▲) and V (△) and protein (○) contents of fractions across the peak are compared. Vanadium was estimated by atomic absorption spectroscopy on wet-ashed samples, while molybdenum was estimated colorimetrically²⁸. Quench corrections were made by the inclusion of internal standards for both metals. Nitrogenase activity was determined by complementation of fractions with Ac2* and assayed as ATP-dependent H₂ evolution as described elsewhere²⁶ (units of activity are nmol H₂ per min).

components as Ac1* for the large protein and Ac2* for the small protein, where the asterisk distinguishes these proteins from the nitrogenase from Mo-sufficient cultures of the parent strain. Some physico chemical properties of the proteins are given in Table 1. Neither component contained significant amounts of Mo. V was present in stoichiometric amounts in Ac1* but not Ac2*. In the final purification step the highest levels of vanadium eluting from the column corresponded with the peak of activity for Ac1* (Fig. 3). Ac2* resembles a typical nitrogenase Fe-protein. Ac1* resembles MoFe-proteins in subunit structure, although one of the two subunits is apparently smaller than that of Ac1 (Fig. 2). The presence of V in Ac1* is unlikely to be adventitious because it is present in stoichiometric amounts in purified protein of high specific activity (Table 1). This finding correlates with our observation that V is required for diazotrophy in MCD1155. Furthermore, our estimate of the specific activity for the vanadium stimulation of nitrogenase in whole cells (120 pmol H₂ evolved per min per pg atoms V) agrees closely with the activity per vanadium in the purified protein (160 pmol H₂ evolved per min per pg atoms V). Compared with H⁺ and N₂ as substrates, C₂H₂ is a poor substrate for the V nitrogenase (Table 1).

Reports that V can substitute for Mo for diazotrophy in azotobacters can be traced to Bortels¹³. Nicholas *et al.*¹⁴ showed that nanomolar levels of V stimulated growth of *A. chroococcum* NCIMB 8003 in a Mo-deficient medium, which agrees with our findings for MCD1155. Becking¹⁵ showed that V replaced Mo in most strains of *A. vinelandii* and *A. chroococcum* tested. Biochemical studies on the vanadium effect led McKenna *et al.*¹⁶ and Burns *et al.*¹⁷ to prepare a crude V-nitrogenase from *A. vinelandii* and *A. chroococcum* which they proposed was the conventional nitrogenase in which Mo had been replaced by V. Some of the properties of their V-nitrogenase (for example, low electron allocation to acetylene) resemble ours. However, their preparations contained significant levels of residual Mo and their findings were attributed to a sparing effect of V on the amounts of Mo required for activity of conventional nitrogenase molybdoprotein¹⁸. Our present findings confirm the reality of a V-nitrogenase but assign it to genetic determinants distinct from those of the Mo-nitrogenase which probably specify the alternative nitrogenase proposed by Bishop and his colleagues³⁻⁵, who suggested⁴ an involvement of V.

We propose in future to refer to conventional nitrogenase as Mo-nitrogenase and the alternative genetically distinct

Table 1 Comparison of physicochemical properties of components of Mo- and V-nitrogenase components of *A. chroococcum*

	M_r	Subunit M_r	Metal content (g atoms mol ⁻¹)		Specific activities (nmol product per min per mg protein)		
			Mo	V	H ₂	NH ₃	C ₂ H ₄
Ac1*	210,000	55,000	<0.107	1.6	1,374	350	516
		50,000					
Ac1	227,000	60,000	1.9	ND	2,138	1,521	1,924
Ac2*	60,000	31,500	0.012	0.04	1,107	507	999
Ac2	64,000	30,000	0.2	ND	1,993	1,361	1,830

For V-nitrogenase components (Ac1* and Ac2*), native and subunit relative molecular masses (M_r s) were determined by thin-layer gel filtration in Sephadex G-200 Superfine and SDS-electrophoresis as described elsewhere²⁶. Metal contents and activities were determined as described for Fig. 3. M_r and metal content of Ac1 and Ac2 are from ref. 9. Nitrogenase activity was measured as described in ref. 26. Ac1* (102 µg) was complemented with Ac2* (1 mg); Ac2* (72 µg) was complemented with Ac1 (0.14 mg). Ac1 (58 µg) was complemented with Ac2 (0.32 mg); Ac2 (86.4 µg) was complemented with Ac1 (0.29 mg). ND, not determined.

nitrogenase as V-nitrogenase, and the component proteins of the V-nitrogenase of *Azotobacter chroococcum* as Acl^{*} and Ac2^{*} using the generally accepted nomenclature of Eady *et al.*¹⁹. *A. chroococcum* contains two unlinked *nifH* genes encoding different Fe-proteins. The second copy (*nifH*^{*}) that we have recently cloned and sequenced is likely to encode Ac2^{*} (ref. 20). We have also shown that a *nifK*-like sequence, which might encode one of the Acl^{*} subunits identified here, is present in the genome of MCD1155 (ref. 8).

Vanadium stimulates N₂ fixation in *Xanthobacter (Mycobacterium) flavus*²¹ and *Clostridium butyricum*²⁹. In *Nostoc muscorum* tungstate- or chromate-resistant mutants became dependent on these ions, respectively, but not molybdate, for N₂ fixation²², a situation which may parallel the apparent dependence of MCD1155 on tungstate. Thus, if azotobacters are not unique in this respect, vanadium may be a key element in the nitrogen cycle especially since it is more abundant in soils and fresh water than is molybdenum²³. Furthermore, it suggests that chemical systems based on vanadium which catalytically reduce N₂ (refs 24, 25) may have more biological significance than has hitherto been recognized.

We thank Simon Harrison for bulk growth of MCD1155, and Beryl Scutt and Brenda Hall for typing the manuscript.

Received 4 March; accepted 13 May 1986.

1. Bortels, H. *Arch. Mikrobiol.* **1**, 333-342 (1930).
2. Burgess, B. K. in *Advances in Nitrogen Fixation Research* (eds Veeger, C. & Newton, W. E.) 103-114 (Nijhoff/Junk, Pudoc, 1984).

3. Bishop, P. E., Jarlenski, D. M. L. & Hetherington, D. R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7342-7346 (1980).
4. Bishop, P. E., Jarlenski, D. M. L. & Hetherington, D. R. *J. Bact.* **150**, 1244-1251 (1982).
5. Bishop, P. E. & Eady, R. R. in *Nitrogen Fixation Research Progress* (eds Evans, H. J., Bottomley, P. J. & Newton, W. E.) 622 (Nijhoff, Dordrecht, 1985).
6. Hales, B. J., Case, E. E. & Langosch, D. in *Nitrogen Fixation Research Progress* (eds Evans, H. J., Bottomley, P. J. & Newton, W. E.) 612 (Nijhoff, Dordrecht, 1985).
7. Chisnell, J. R. & Bishop, P. E. in *Nitrogen Fixation Research Progress* (eds Evans, H. J., Bottomley, P. J. & Newton, W. E.) 623 (Nijhoff, Dordrecht, 1985).
8. Robson, R. L. *Archiv. Microbiol.* (in the press).
9. Yates, M. G. & Planqué, K. *Eur. J. Biochem.* **60**, 467-476 (1975).
10. Jones, R., Woodley, P. R. & Robson, R. L. *Molec. gen. Genet.* **197**, 318-327 (1984).
11. Takahashi, H. & Nason, A. *Biochim. biophys. Acta* **23**, 433-435 (1957).
12. Keeler, R. F. & Varner, J. E. *Archs Biochem. Biophys.* **70**, 585-590 (1957).
13. Bortels, H. *Zentre. Bakteriell. Parasitenk. Abt. II* **95**, 193-218 (1936).
14. Nicholas, D. J. D., Fisher, D. J., Redmond, W. J. & Wright, M. A. *J. gen. Microbiol.* **22**, 191-205 (1960).
15. Becking, J. H. *Pl. Soil* **2**, 171-201 (1962).
16. McKenna, C. E., Benemann, J. R. & Traylor, T. C. *Biochem. biophys. Res. Commun.* **41**, 1501-1508 (1970).
17. Burns, R. C., Fuchsman, W. H. & Hardy, R. W. F. *Biochem. biophys. Res. Commun.* **42**, 353-358 (1971).
18. Benemann, J. R., McKenna, C. E., Lie, R. F., Traylor, T. G. & Kamen, M. D. *Biochim. biophys. Acta* **264**, 25-38 (1972).
19. Eady, R. R., Smith, B. E., Cook, K. A. & Postgate, J. R. *Biochem. J.* **128**, 655-675 (1972).
20. Robson, R., Woodley, P. & Jones, R. *EMBO J.* **5**, 1159-1163.
21. Krylova, N. B. *Dokl. Mosk. Sel'Kokhoz. Akad.* **84**, 293-299 (1963).
22. Singh, N. H., Vaishampayan, A. & Singh, R. K. *Biochem. biophys. Res. Commun.* **81**, 67-74 (1978).
23. Bowen, H. J. M. *Trace Elements in Biochemistry* (Academic, London, 1966).
24. Denisov, N. T., Efimov, O. N., Shuvalova, N. I., Shilova, A. K. & Shilov, A. E. *Russ. J. phys. Chem.* **44**, 1693-1694 (1970).
25. Nikinova, L. A., Ovcharenko, A. G., Efimov, O. N., Avilov, V. A. & Shilov, A. E. *Kin. i. Kat.* **13**, 1602-1603 (1972).
26. Eady, R. R. *Meth. Enzym.* **69**, 751-776 (1980).
27. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
28. Bulen, W. A. & Le Comte, J. R. *Proc. natn. Acad. Sci. U.S.A.* **51**, 979-989 (1966).
29. Jensen, H. L. & Spencer, D. *Proc. Linn. Soc. N. S. Wales* **72**, 73-86.

Botulinum C2 toxin ADP-ribosylates actin

K. Aktories*, M. Bärmann†, I. Ohishi‡,
S. Tsuyama‡, K. H. Jakobs§ & E. Habermann*

* Rudolf-Buchheim-Institut für Pharmakologie der Universität
Giessen, Frankfurter Strasse 107, D-6300 Giessen, FRG

† Institut für Pharmakologie und Toxikologie der Universität Giessen,
D-6300 Giessen, FRG

‡ College of Agriculture, University of Osaka Prefecture, Sakai,
Osaka 591, Japan

§ Pharmakologisches Institut der Universität Heidelberg, D-6900
Heidelberg, FRG

ADP-ribosylation of regulatory proteins is an important pathological mechanism by which various bacterial toxins affect eukaryotic cell functions. While diphtheria toxin catalyses the ADP-ribosylation of elongation factor 2, which results in inhibition of protein synthesis, cholera toxin and pertussis toxin ADP-ribosylate N_s and N_i, respectively, the GTP-binding regulatory components of the adenylate cyclase system, thereby modulating the bidirectional hormonal regulation of the adenylate cyclase^{1,2}. Botulinum C2 toxin is another toxin which has been reported to possess ADP-ribosyltransferase activity³. This extremely toxic agent is produced by certain strains of *Clostridium botulinum*⁴ and induces hypotension⁵, an increase in intestinal secretion⁶, vascular permeability⁷ and haemorrhaging in the lungs⁵. In contrast to botulinum neurotoxins, the botulinum C2 toxin apparently lacks any neurotoxic effects⁵. Here we report that botulinum C2 toxin ADP-ribosylates a protein of relative molecular mass 43,000 (43K) in intact cells and in cell-free preparations. We present evidence that the 43K protein substrate is actin, which is apparently mono-ADP-ribosylated by the toxin. Botulinum C2 toxin also ADP-ribosylated purified liver G-actin, whereas liver F-actin was only poorly ADP-ribosylated and skeletal muscle actin was not ADP-ribosylated in

either its G form or its F form. ADP-ribosylation of liver G-actin by botulinum C2 toxin resulted in a drastic reduction in viscosity of actin polymerized *in vitro*.

Botulinum C2 toxin is a binary toxin and consists of two components, C2-I and C2-II, with relative molecular masses (*M_r*s) of ~50K and 100K, respectively⁴. Both components are actually completely separate proteins which interact to cause the toxic effects. The 50K component, which is non-toxic in intact tissue, apparently represents the ADP-ribosyltransferase³. The 100K component appears to be involved in the binding of the toxin to intact cells⁶ and probably enables the enzymatically active component C2-I to enter the cell. Therefore, we used the combination of C2-I and -II for studies in intact cells and the isolated C2-I component for studies in cell-free preparations. In platelet lysates, C2-I toxin ADP-ribosylated a 43K protein, predominantly located in the cytosolic fraction (Fig. 1a). ADP-ribosylation of a 43K protein was also found in lysates of S49 lymphoma cells, hamster adipocytes, chicken embryo fibroblasts and neuroblastoma x glioma hybrid cells (NG 108-15) and in brain homogenates (not shown). As botulinum C2-I toxin ADP-ribosylated a protein with a *M_r* similar to those of the N_s and N_i proteins, the substrates of cholera and pertussis toxin, respectively, we studied the ADP-ribosylated substrates of all three toxins by SDS-gel electrophoresis. As shown in Fig. 1b, the 43K substrate of botulinum C2 toxin was clearly separated from the cholera toxin and pertussis toxin ADP-ribosylated regulatory components, N_s and N_i, respectively, of the adenylate cyclase system. When intact S49 lymphoma cells were pretreated with botulinum C2 toxin (components I and II), the subsequent C2-I-induced ADP-ribosylation of the 43K substrate protein in cell-free preparations was largely decreased (Fig. 2a). These findings suggest that the substrate of botulinum C2 toxin is identical in intact cells and in cell-free preparations. C2-I-induced ADP-ribosylation of the 43K protein in platelet cytosol was time- and concentration-dependent. In the presence of 5 μM NAD, maximal ADP-ribosylation was seen at ~1 μg ml⁻¹ toxin after 2 hours of incubation. That the ADP-ribosylation of the 43K protein was a mono-ADP-ribosylation under the conditions

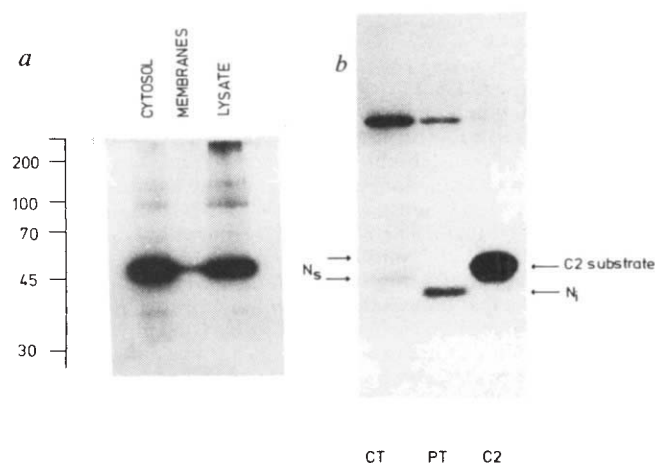


Fig. 1 Polyacrylamide gel analysis of radioactive products resulting from ADP-ribosylation catalysed by botulinum C2-I toxin, cholera toxin and pertussis toxin. *a*, ADP-ribosylation of a 43K protein in human platelet lysate, crude platelet membranes and platelet cytosol by botulinum C2-I toxin. *b*, ADP-ribosylation of the N_s and N_c proteins, and of the 43K C2 substrate in crude membranes of S49 lymphoma cells by cholera toxin (CT), pertussis toxin (PT) and botulinum C2-I toxin (C2), respectively.

Methods. *a*, Human platelets were isolated as described previously¹¹, then lysed by freezing and thawing in a hypotonic medium containing 10 mM triethanolamine-HCl (pH 7.4) and 5 mM EDTA. Platelet cytosol was obtained by centrifuging the lysate for 15 min at 30,000g. The pellet was washed twice with the hypotonic buffer then used as the crude platelet membrane fraction. ADP-ribosylation was carried out in a buffer containing 10 mM thymidine, 5 mM $MgCl_2$, 1 mM EDTA, $1 \mu g ml^{-1}$ C2-I toxin, $0.5 \mu M$ [α - ^{32}P] NAD ($\sim 600,000$ c.p.m.), 0.5 mM ATP and 50 mM triethanolamine-HCl (pH 7.4). Protein concentration was 50–200 μg per tube. After incubation of 37 °C for 1 h, the reaction was stopped by adding 1 ml of trichloroacetic acid (20% w/v). The resulting pellet was washed with ether and dissolved in 50 μl of electrophoresis buffer. SDS-gel electrophoresis was performed according to Laemmli¹². Gels were stained with Coomassie blue, destained and subjected to autoradiography for 24–48 h. Thereafter, the protein bands containing the radioactive label were cut from destained gels, solubilized in the presence of 1 ml of 30% hydrogen peroxide and counted for ^{32}P . ADP-ribosylation of the 43K protein in platelet lysate, crude membranes and cytosol was 24.7, 7.6 and 28.3 pmol ADP-ribose per mg protein per h, respectively. *b*, Crude S49 lymphoma cell membranes, prepared essentially as described elsewhere¹³, were incubated with 100 $\mu g ml^{-1}$ cholera toxin, 40 $\mu g ml^{-1}$ pertussis toxin or 1 $\mu g ml^{-1}$ C2-I toxin for 45 min at 37 °C in medium containing 1 μM [α - ^{32}P] NAD ($\sim 1 \mu Ci$ per tube), 10 mM thymidine, 10 mM arginine, 5 mM creatine phosphate, 0.1 mg ml^{-1} creatine kinase, 5 mM $MgCl_2$, 1 mM EDTA, 0.5 mM ATP, 0.5 mM GTP and 50 mM triethanolamine-HCl (pH 7.4). Cholera toxin and pertussis toxin were preactivated by incubation in the presence of 20 mM dithiothreitol (DTT) for 10 min at 37 °C.

used is indicated by the following observations. First, whereas the total amount of the ADP-ribosylation increased with length of incubation, the M_r of the labelled substrate remained constant. Second, addition of snake venom phosphodiesterase hydrolysed the protein-bound radioactivity and resulted in the formation of 5'AMP (not shown).

The similarity between the relative molecular mass of the labelled protein and that of actin led us to study whether the platelet microfilament protein was ADP-ribosylated by botulinum C2-I toxin. After incubation with anti-actin antibodies but not with non-immune rabbit serum, protein A-agarose precipitated the toxin-labelled platelet cytosolic protein (Fig. 3a). In addition, immunoblotting with anti-actin antibodies clearly demonstrated that the antibodies recognized the labelled 43K protein (Fig. 3b). Both findings strongly indicated that actin

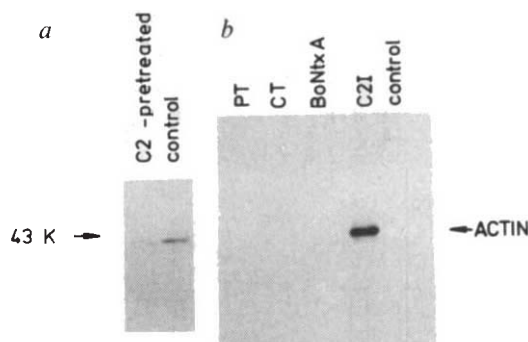


Fig. 2 *a*, Effect of pretreatment of intact S49 lymphoma cells with botulinum C2 toxin on the subsequent ADP-ribosylation of cell lysates. *b*, Effects of botulinum C2-I toxin, cholera toxin, pertussis toxin and botulinum neurotoxin A on the radiolabelling of liver G-actin.

Methods. *a*, S49 lymphoma cells were treated for 24 h without (control) or with botulinum C2 toxin ($5 \mu g l^{-1}$ component I plus $10 \mu g l^{-1}$ component II), then crude membranes were prepared and treated with botulinum C2-I toxin ($1 \mu g ml^{-1}$) in the presence of [α - ^{32}P]NAD. SDS-gel electrophoresis and autoradiography were performed as described in Fig. 1 legend. *b*, Purified liver G-actin ($100 \mu g ml^{-1}$), prepared according to Jaberg¹⁴, was incubated without (control) or with 1 $\mu g ml^{-1}$ C2-I toxin (C21), 200 $\mu g ml^{-1}$ preactivated cholera toxin (CT), 30 $\mu g ml^{-1}$ preactivated pertussis toxin (PT) or 80 $\mu g ml^{-1}$ botulinum neurotoxin A (BoNtx A) in the presence of $0.5 \mu M$ [α - ^{32}P]NAD ($\sim 500,000$ c.p.m. per tube) as described in Fig. 1 legend.

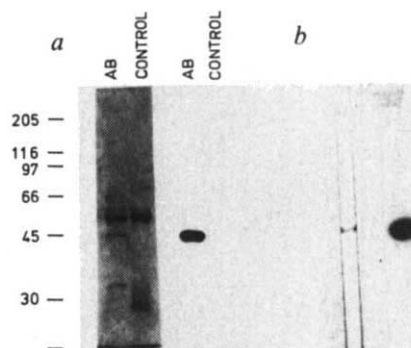


Fig. 3 Immunoprecipitation and immunoblotting of the platelet 43 K protein labelled by botulinum C2-I toxin. *a*, SDS-gel electrophoresis and autoradiography of the ADP-ribosylated C2-I toxin substrate immunoprecipitated by anti-actin antibodies (AB) but not by non-immune serum (Control). *b*, Immunoblot and autoradiography of the 43 K platelet cytosolic protein ADP-ribosylated by C2 toxin.

Methods. *a*, Platelet cytosol ($\sim 200 \mu g$ protein) was ADP-ribosylated as described in Fig. 1 legend, except that NAD and $MgCl_2$ was at 1 μM and 0.5 mM, respectively and the total volume was 0.5 ml. Toxin-ADP-ribosylated platelet cytosol (200 μl), freed from unreacted [α - ^{32}P]NAD by gel-filtration, was incubated in 1% Triton X-100, 1% deoxycholate, 0.1% SDS, 150 mM NaCl and 50 mM Tris-HCl (pH 7.2) with 100 μl of rabbit polyclonal antibodies ($0.1 mg ml^{-1}$) for 12 h at 0 °C in a total volume of 400 μl . The antibodies were raised against heat- and SDS-denatured pig liver actin and affinity-purified over a column of rabbit skeletal muscle actin coupled to Sepharose 4B. Then, 20 μl of protein A-agarose (Sigma) was added for 1 h. The immuno-protein-A-agarose complex was separated by centrifugation, washed twice and analysed by SDS-gel electrophoresis. *b*, Anti-actin antibodies recognized the ADP-ribosylated toxin substrate. Immunoblotting was performed according to Towbin *et al.*¹⁵ with peroxidase-coupled swine IgG to rabbit IgG (Dakopatts) as second antibody and 4-chloro-1-naphthol as peroxidase substrate.

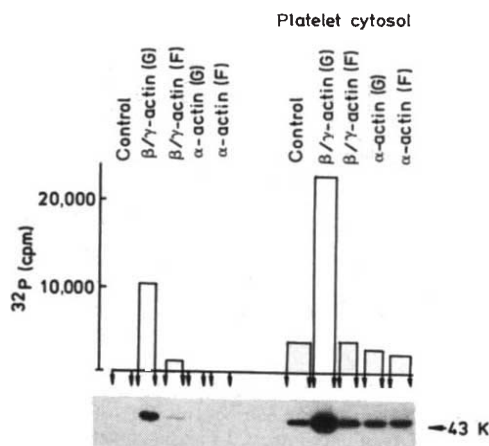


Fig. 4 Botulinum C2-I toxin-induced ADP-ribosylation of actin. For ADP-ribosylation purified liver G-actin ($200 \mu\text{g ml}^{-1}$) and F-actin ($160 \mu\text{g ml}^{-1}$) and rabbit muscle G-actin ($200 \mu\text{g ml}^{-1}$) and F-actin ($110 \mu\text{g ml}^{-1}$), prepared according to Pardee and Spudich¹⁶, were incubated without or with human platelet cytosol ($170 \mu\text{g ml}^{-1}$) in the presence of $1 \mu\text{g ml}^{-1}$ botulinum C2-I toxin and $0.2 \mu\text{M}$ ^{32}P -NAD ($\sim 250,000$ c.p.m. per tube) for 15 min at 37°C . Reaction medium, SDS-gel electrophoresis, autoradiography and ^{32}P measurements were as described in Fig. 1 legend.

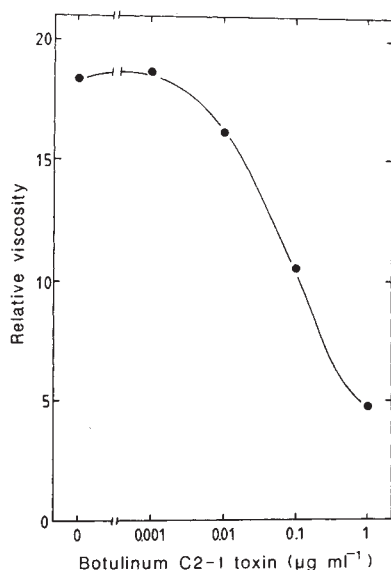


Fig. 5 Effect of C2 toxin-induced ADP-ribosylation on β/γ -actin polymerization, measured as relative viscosity.

Methods. Liver G-actin ($30 \mu\text{M}$) was incubated in the presence of 1 mM NAD, 10 mM thymidine, 0.5 mM MgCl_2 , $0.5 \mu\text{M}$ ATP and 50 mM triethanolamine-HCl (pH 7.4) without or with botulinum C2-I toxin at the indicated concentrations for 1 h at 37°C . Thereafter, the samples were placed on ice. Polymerization was initiated in $140 \mu\text{l}$ of the pretreated actin solution by adding $20 \mu\text{l}$ of a medium containing 16 mM MgCl_2 , 1.6 mM CaCl_2 , 4.0 mM ATP, 0.8 mM DTT and 16 mM Tris-HCl, pH 7.5. Measurements were made after 10 min at 25°C . Viscosity was determined using the falling-ball device described by Pollard and Cooper¹⁷. Relative viscosity is the quotient sample viscosity (η_s) divided by the buffer viscosity (η_a).

was the protein ADP-ribosylated by botulinum C2 toxin. Furthermore, purified liver G-actin was a good substrate for botulinum C2-I toxin (Fig. 2b). In contrast, cholera toxin and pertussis toxin did not ADP-ribosylate the isolated microfilament protein. Botulinum neurotoxin A, which has no enterotoxigenic effects⁸, showed no ADP-ribosyltransferase activity with platelet lysates (not shown), brain homogenates (not shown), with purified liver G-actin. Interestingly, only the G-form of liver actin was a good substrate for the ADP-ribosylation by botulinum C2-I toxin. The F-form of liver actin was much less ADP-ribosylated, and purified α -actin from rabbit skeletal muscle was not ADP-ribosylated in its G- or F-form (Fig. 4). Thus, it seems that botulinum C2 toxin-induced ADP-ribosylation is highly specific for β/γ -actin in its G-form. The slight ADP-ribosylation of liver F-actin may be due to a minor fraction of G-actin present in the F-actin preparation. Accordingly, we found that phalloidin, which specifically binds to F-actin and which stabilizes this form⁹, impaired or abolished the toxin-induced ADP-ribosylation (data not shown). The addition of platelet cytosol to G-actin further increased the amount of ADP-ribosylation, suggesting that an additional factor present in the cytosolic fraction facilitated the ADP-ribosylation reaction. In contrast, the addition of platelet cytosol did not increase the ADP-ribosylation of liver F-actin or muscle α -actin.

In order to study whether the ADP-ribosylation of actin caused any change in the functional properties of the microfilament protein, we compared the viscosity of control and C2-I toxin-pretreated actin 10 min after induction of polymerization of actin with MgCl_2 . When actin was pretreated with increasing concentrations of botulinum C2-I toxin, the relative viscosity of the polymerized actin solution was decreased in a concentration-dependent manner (Fig. 5). At $1 \mu\text{g ml}^{-1}$ C2-I toxin, the relative viscosity of ADP-ribosylated actin was reduced by about 72% compared with the control preparation, indicating that the polymerization of the modified actin was largely impaired.

Taken together, the data presented show that botulinum C2 toxin induces the ADP-ribosylation of β/γ -actin. This covalent modification subsequently alters at least one functional property of the microfilament protein, actin polymerization. It is suggested that the ADP-ribosylation of actin is the pathophysiological basis of the toxic effects of botulinum C2 toxin, for example, the increase in vascular permeability, gut secretion and haemorrhaging of the lungs. Recent findings that the C2 toxin induces rounding up and subsequent lysis of cultured cells, can also be interpreted in this sense¹⁰. Botulinum C2 toxin, which clearly discriminates between non-muscle and skeletal muscle actin, appears to be an important tool for studying the physiological processes which involve non-muscle actin.

Received 21 January; accepted 7 May 1986.

1. Foster, J. W. & Kinney, D. M. *CRC crit. Rev. Microbiol.* **11**, 273-297 (1985).
2. Gill, D. M. & Woolkalis, M. *Ciba Fdn Symp.* **112**, 47-69 (1985).
3. Simpson, L. L. *J. Pharmac. exp. Ther.* **230**, 665-669 (1984).
4. Ohishi, I., Iwasaki, M. & Sakaguchi, G. *Infect. Immunity* **30**, 668-673 (1980).
5. Simpson, L. L. *J. Pharmac. exp. Ther.* **223**, 695-701 (1982).
6. Ohishi, I. *Infect. Immunity* **40**, 691-695 (1983).
7. Ohishi, I. *Infect. Immunity* **40**, 336-339 (1983).
8. Simpson, L. L. *Pharmac. Rev.* **33**, 155-188 (1981).
9. Wieland, T. *Naturwissenschaften* **64**, 303-309 (1977).
10. Ohishi, I., Miyake, M., Ogura, H. & Nakamura, S. *FEMS Microbiol. Lett.* **23**, 281-284 (1984).
11. Jakobs, K. H., Saur, W. & Schultz, G. *Naunyn Schmiedeberg's Arch. Pharmac.* **302**, 285-291 (1978).
12. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
13. Jakobs, K. H., Gehring, U., Gaugler, B., Pfeuffer, T. & Schultz, G. *Eur. J. Biochem.* **130**, 605-611 (1983).
14. Jäberg, B. Z. *Naturforsch.* **38c**, 829-833 (1983).
15. Towbin, H., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350-4354 (1979).
16. Pardee, J. D. & Spudich, J. A. *Meth. Enzym.* **85**, 164-181 (1982).
17. Pollard, T. D. & Cooper, J. A. *Meth. Enzym.* **85**, 211-230 (1982).